

On the Existence of Characterization Logics and Fundamental Properties of Argumentation Semantics

Der Fakultät für Mathematik und Informatik
der Universität Leipzig
eingereichte

Habilitationsschrift

zur Erlangung des akademischen Grades

doctor rerum naturalium habilitatus
(Dr. rer. nat. habil.)

vorgelegt

von Dr. rer. nat. Ringo Baumann

geboren am 05.09.1980 in Hohenmölsen

Die Annahme der Habilitationsschrift haben empfohlen:

- | | |
|---|---|
| 1. Professor Dr. Gerhard Brewka, | Universität Leipzig, Deutschland |
| 2. Professor Dr. Dr. Dr. Dov Gabbay, | King's College London, Vereinigtes Königreich |
| 3. Professor Dr. Mirosław Truszczyński, | University of Kentucky, Vereinigte Staaten |

Beschluß über die Verleihung des akademischen Grades vom 18.11.2019.

Preface

Acknowledgements	v
Abstract	vii
Kurzfassung	ix
Contents	xix
List of Figures	xxi

Acknowledgements

Many factors have contributed in one way or another to the successful completion of this thesis. Among them are: scientific challenges, an inspiring working atmosphere, thinking, asking questions, fruitful discussions with colleagues, research stays, listening to and giving talks at conferences, writing papers, getting them accepted or rejected, comments of anonymous reviewers, giving lessons, answering student questions, supervision of degree theses, watching and doing sports, listening to and making music and of course, spending time with friends and family.

I was lucky to meet and get in contact with many inspiring colleagues. Some of them deserve a separate mentioning as they influenced my own research in a variety of ways: Pietro Baroni, Gerd Brewka, Wolfgang Dvořák, Dov Gabbay, Heinrich Herre, Thomas Linsbichler, Frank Loebe, Christof Spanring, Hannes Straß and Stefan Woltran.

Abstract

Given the large variety of existing logical formalisms it is of utmost importance to select the most adequate one for a specific purpose, e.g. for representing the knowledge relevant for a particular application or for using the formalism as a modeling tool for problem solving. Awareness of the nature of a logical formalism, in other words, of its fundamental intrinsic properties, is indispensable and provides the basis of an informed choice.

One such intrinsic property of logic-based knowledge representation languages is the context-dependency of pieces of knowledge. In classical propositional logic, for example, there is no such context-dependence: whenever two sets of formulas are equivalent in the sense of having the same models (*ordinary equivalence*), then they are mutually replaceable in arbitrary contexts (*strong equivalence*). However, a large number of commonly used formalisms are not like classical logic which leads to a series of interesting developments. It turned out that sometimes, to characterize strong equivalence in formalism \mathcal{L} , we can use ordinary equivalence in formalism \mathcal{L}' : for example, strong equivalence in normal logic programs under stable models can be characterized by the standard semantics of the logic of here-and-there. Such results about the existence of *characterizing logics* has rightly been recognized as important for the study of concrete knowledge representation formalisms and raise a fundamental question: *Does every formalism have one?* In this thesis, we answer this question with a qualified “yes”. More precisely, we show that the important case of considering only finite knowledge bases guarantees the existence of a canonical characterizing formalism. Furthermore, we argue that those characterizing formalisms can be seen as *classical*, monotonic logics which are uniquely determined (up to isomorphism) regarding their model theory.

The other main part of this thesis is devoted to argumentation semantics which play the flagship role in Dung’s abstract argumentation theory. Almost all of them are motivated by an easily understandable intuition of what should be acceptable in the light of conflicts. However, although these intuitions equip us with short and comprehensible formal definitions it turned out that their intrinsic properties such as *existence and uniqueness*, *expressibility*, *replaceability* and *verifiability* are not that easily accessible. We review the mentioned properties for almost all semantics available in the literature. In doing so we include two main axes: namely first, the distinction between extension-based and labelling-based versions and secondly, the distinction of different kind of argumentation frameworks such as finite or unrestricted ones.

Kurzfassung

Hintergrund

Die *Künstliche Intelligenz (KI)* ist ein Teilgebiet der Informatik welches sich grob gesagt mit der Automatisierung von intelligenten Verhalten beschäftigt. Allwissende Frage-Antwort Systeme wie IBMs *Watson*¹ als auch DeepMinds *AlphaZero*², welches in übermenschlicher Stärke Schach, Shogi als auch Go spielen kann, haben zu einer verstärkten Wahrnehmung der KI in der breiten Öffentlichkeit gesorgt. Insbesondere das zuletzt erwähnte Programm wird von nicht wenigen Forschern als der größte Durchbruch der KI in der letzten Dekade angesehen. Selbst die optimistischsten Wissenschaftler haben nicht daran geglaubt, daß ein Programm die menschliche Elite im Gospiel in der näheren Zukunft schlagen könne, da die Komplexität dieses Spiels so unglaublich groß ist, unvergleichbar größer als beim Schach. Der dänische Schachgroßmeister Peter Heine Nielsen³ sagte über die Schachfähigkeiten von AlphaZero das Folgende:

I always wondered how it would be if a superior species landed on earth and showed us how they played chess. Now I know.

Ich habe mich oft gefragt, wie es wohl wäre, wenn eine überlegene Spezies zur Erde käme und mit uns Schach spiele. Jetzt weiß ich es.

Wie erreicht man solch eine enorme Spielstärke? Alpha Zero arbeitet mit einem speziellen *neuronalen Netz*, welches Zugvorschläge als auch Stellungsbewertungen liefert. Künstliche neuronale Netze sind simplifizierte Nachbildungen von neuronalen Strukturen wie sie in der Natur vorkommen. In einer Trainingsphase wird ein solches System angelernt und extrahiert daraus eine Bewertungsfunktion, welche es für zukünftige Entscheidungen verwendet. Das bemerkenswerte an AlphaZero ist, daß es nicht nur ein Spezialist für ein einziges Spiel ist. Des Weiteren hatte es keinen Zugriff auf Eröffnungstabellen oder gar Schachbibliotheken und lernte somit gängige Eröffnungsvarianten und Spielstrategien nur anhand der Spielregeln und eigenen Trainingspartien. AlphaZero zeigt eindrucksvoll, daß künstliche neuronale Netze eine erfolgsversprechende Möglichkeit sind, Wissen zu repräsentieren und zu verarbeiten. Die Fragestellung, wie Wissen intern dargestellt und weiterverarbeitet werden soll, ist ein entscheidender Punkt für jeden Schaffer einer künstlichen Intelligenz. Die Beantwortung eben dieser, hängt stark von den ausgerufenen Zielen und dem Anwendungsgebiet im Allgemeinen ab. In Anbetracht des breiten Spektrums an existierenden Wissensrepräsentationsformalismen ist es unabdingbar, profunde Kenntnisse über das Wesen eines logischen Formalismus, mit anderen Worten, seiner grundlegenden intrinsischen Eigenschaften zu haben, um nachvollziehbare und bewußte Entscheidungen zu fällen. Neben dem tieferen

¹Ferrucci et al., 2010, *Building Watson: An Overview of the DeepQA Project*, AI Magazine, 31(3):59–79.

²Silver et al., 2018, *A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play*, Science, 362(6419):1140–1144.

³BBC, 2017, *Google's 'superhuman' DeepMind AI claims chess crown*, <https://www.bbc.com/news/technology-42251535>.

Verständnis des betrachteten Formalismus, kann das Studium solcher inhärenten Eigenschaften dazu beitragen, interessante Fragmente zu identifizieren oder auch nützliche Erweiterungen eines Formalismus zu entwickeln. Darüber hinaus können die gewonnenen Erkenntnisse verwendet werden, um vorhandene Algorithmen zu verfeinern oder vollständig neue zu entwerfen.

Die vermutlich bekannteste intrinsische Eigenschaft eines logischen Formalismus ist die *Monotonie*. Monotone Logiken, wie zum Beispiel die klassische Aussagenlogik oder die Prädikatenlogik 1. Stufe, eignen sich hervorragend für die Formalisierung universeller Wahrheiten. Denn Monotonie eines logischen Formalismus bedeutet nicht weniger als, daß wenn eine Formel ϕ eine Folgerung aus einer Theorie T ist, diese für immer und ausnahmslos erhalten bleibt, egal wieviel neues Wissen wir zur Ausgangstheorie T hinzufügen. Nichtmonotone Logiken⁴, also Logiken, welche die Monotonieeigenschaft verletzen, ermöglichen es, anfechtbare Schlußfolgerungen zu ziehen. Das bedeutet, daß eine solche Schlußfolgerung gegebenenfalls zu einem späteren Zeitpunkt wieder verworfen wird. Beide Arten von Logiken haben ihre traditionellen Anwendungsdomänen wie zum Beispiel das Schließen in der Mathematik, in welcher man keinesfalls einmal formal bewiesene Eigenschaften wieder verwerfen will oder im Gegensatz dazu, die Formalisierung des Alltagsschließens, welche ohne die Rücknahme von auf unvollständigen Wissen basierenden Schlüssen nicht vorstellbar wäre. Abgesehen von dieser grundlegenden Wahlmöglichkeit bezüglich der Monotonie gibt es viele weitere Vergleichskriterien, die die Entscheidung der zu wählenden Logik beeinflussen.

Eine dieser weiteren entscheidenden Eigenschaften ist die sogenannte *Berechnungskomplexität*.⁵ Die Komplexität eines algorithmisch lösbaren Entscheidungsproblems beziehungsweise die Zugehörigkeit von eben diesen zu einer bestimmten Komplexitätsklasse kann als Hinweis darauf verstanden werden, ob ein vorgegebenes Problem praktisch, das heißt mit einem vertretbaren zeitlichen Aufwand lösbar ist. Ein weiterer wichtiger Sachverhalt ist die sogenannte *Modularität*. Hierbei beschäftigt man sich unter anderem mit der Frage, ob es möglich ist, eine gegebene Theorie in Untertheorien so aufzuteilen, daß die formale Semantik der Ausgangstheorie durch die Semantik der Untertheorien rekonstruiert werden kann. Beide genannten Eigenschaften wurden eingehend für führende nicht-monotone Formalismen wie etwa *Standardannahmelogik*^{6,7} als auch *abstrakte Argumentationsrahmenwerke*^{8,9} unter verschiedenen Argumentationssemantiken studiert.

Ein Hauptteil dieser Habilitationsschrift widmet sich der Untersuchung von vier weiteren grundlegenden intrinsischen Eigenschaften von Semantiken abstrakter Argumentationsrahmenwerke. Was genau sind solche Rahmenwerke? Abstrakte Argumentationsrahmenwerke wurden 1995 von Phan Minh Dung eingeführt und sind formal betrachtet nichts weiter als gerichtete Graphen.¹⁰ Die Knoten eines gegebenen Graphens werden als Argumente und die gerichteten Kanten als Attacken zwischen solchen interpretiert. Beide Entitäten werden als primitive Elemente behandelt, das heißt weder die interne Struktur von Argumenten noch die Gründe warum ein Argument ein anderes attackiert werden berücksichtigt. Der Hauptfokus von Dungs Argumentationstheorie liegt im Auflösen von Konflikten. Zu diesem Zwecke wurde eine Vielzahl von Semantiken eingeführt, wobei jede von ihnen auf eine bestimmte Art und Weise akzeptable Mengen von Argumenten definiert, sogenannte *Erweiterungen*. Ein zweiter weit verbreiteter Ansatz zum Auflösen von Konflikten sind sogenannte *kennzeichnungsbasierte* Semantiken. Derlei Semantiken

⁴Gerhard Brewka, 1989, *Nonmonotonic logics – A brief overview*, AI Communications, 2(2):88–97.

⁵Christos H. Papadimitriou, 1994, *Computational complexity*, Addison-Wesley, I-XV, 1-523.

⁶Georg Gottlob, 1992, *Complexity results for nonmonotonic logics*, Journal of Logic and Computation, 2(3):397–425.

⁷Hudson Turner, 1996, *Splitting a default theory*, IAAI, 645–651.

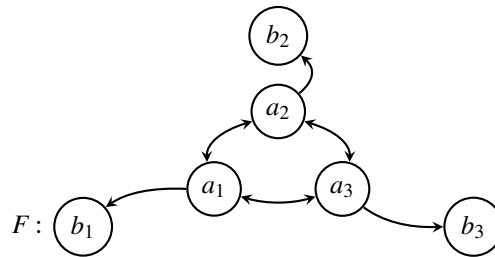
⁸Wolfgang Dvorák und Paul E. Dunne, 2018, *Computational problems in formal argumentation and their complexity*, Handbook of Formal Argumentation, Kapitel 14.

⁹Baroni et al., 2018, *Abstract argumentation frameworks and their semantics*, Handbook of Formal Argumentation, Kapitel 4.

¹⁰Phan Minh Dung, 1995, *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games*, Artificial Intelligence, 77:321–357.

enthalten mehr Informationen als ihre erweiterungsbasierten Gegenstücke. Kennzeichnungsbasierte Semantiken geben nicht nur an, welche Argumente explizit akzeptiert sind, also in einer Erweiterung enthalten, sondern auch, welche Argumente explizit verworfen sind, das heißt von akzeptierten Argumenten attackiert, und auch jene, welche weder akzeptiert, noch verworfen sind, sogenannte unentschlossene Argumente.

Der nachfolgende gerichtete Graph F ist ein einfaches Beispiel eines abstrakten Argumentationsrahmenwerkes, welches aus sechs Argumenten besteht. Eine der bekanntesten Auswertungsfunktionen, nämlich die erweiterungsbasierte Version der *stabilen* Semantik würde die Mengen $\{a_1, b_2, b_3\}$, $\{a_2, b_1, b_3\}$ und auch $\{a_3, b_1, b_2\}$ als akzeptable Mengen, sprich stabile Erweiterungen ausgeben. Jede einzelne stabile Erweiterung trägt keine Konflikte, das heißt akzeptierte Argumente attackieren sich nicht und des Weiteren, werden alle übrigen Argumente widerlegt, soll heißen, sie werden von mindestens einem akzeptierten Argument attackiert.



Fast alle Semantiken sind motiviert durch leicht verständliche Anschauungen darüber, welche Argumente im Beisein von Konflikten kollektiv akzeptabel sein sollen. Obwohl diese Intuitionen zu kurzen und formal unaufwendigen Definitionen führen, stellt sich heraus, daß grundlegende Eigenschaften wie *Existenz und Eindeutigkeit*, *Ausdruckskraft*, *Ersetzbarkeit* als auch *Überprüfbarkeit* nicht so leicht zugänglich beziehungsweise faßbar sind. Im Folgenden werden wir diese Eigenschaften genauer spezifizieren. Bei der Untersuchung dieser Eigenschaften wurden zwei Hauptachsen mit eingeschlossen: Erstens, die Unterscheidung zwischen erweiterungsbasierten und kennzeichnungsbasierten Versionen und zweitens, die Unterscheidung verschiedener syntaktischer Klassen von abstrakten Argumentationsrahmenwerken, wie zum Beispiel, Rahmenwerke mit nur endlich vielen und jene mit unbeschränkt vielen Argumenten.

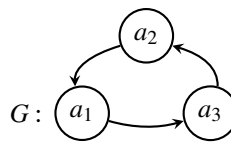
1. *Existenz und Eindeutigkeit* Ist es möglich, und wenn ja wie, die Existenz von mindestens einer beziehungsweise genau einer Erweiterung/Kennzeichnung zu garantieren?
2. *Ausdruckskraft* Ist es möglich, und wenn ja wie, eine bestimmte vorgegebene Menge von Erweiterungen/Kennzeichnungen mit Hilfe eines einzelnen Argumentationsrahmenwerkes zu realisieren?
3. *Ersetzbarkeit* Ist es möglich, und wenn ja wie, bestimmte Teile eines Rahmenwerkes so zu vereinfachen, daß die modifizierte Version von dem Ausgangsrahmenwerk nie durch gleichzeitiges Hinzufügen von neuer Information semantisch unterschieden werden kann?
4. *Überprüfbarkeit* Ist es möglich, und wenn ja wie, mit strikt weniger Information, als jene formal gegeben durch ein abstraktes Argumentationsrahmenwerk, die Semantik eines eben solchen eindeutig zu berechnen?

Existenz und Eindeutigkeit

Gegeben ein beliebiger logischer Formalismus \mathcal{L} zusammen mit einer Semantik $\sigma_{\mathcal{L}}$. Eine zentrale Frage in Bezug auf seine Eignung für konkrete Anwendungen ist, ob die Semantik einer jeden \mathcal{L} -Theorie T eine formale Bedeutung zuweist, formal ausgedrückt $|\sigma_{\mathcal{L}}(T)| \geq 1$. Eine stärkere Forderung als die Existenz einer formalen Bedeutung ist die Eindeutigkeit, das heißt $|\sigma_{\mathcal{L}}(T)| = 1$ für jede \mathcal{L} -Theorie T . Natürlich sind diese Eigenschaften aus verschiedenen Perspektiven interessant. Im Falle der eindeutigen Bestimmtheit beobachten wir das Zusammenfallen von der skeptischen und leichtgläubigen Schlußweise. Genauer gesagt, sofern $\sigma_{\mathcal{L}}(T) = \{E\}$, dann $\bigcap \sigma_{\mathcal{L}}(T) = \bigcup \sigma_{\mathcal{L}}(T) = E$. Sollte für eine spezifische Anwendung eine \mathcal{L} -Theorie T nur dann als bedeutungsvoll interpretiert werden, sofern $\sigma_{\mathcal{L}}(T) \neq \emptyset$, dann wäre die Existenz eine erstrebenswerte Eigenschaft. Wenn dies im Allgemeinen für eine Semantik $\sigma_{\mathcal{L}}$ nicht gegeben ist, kann man versuchen, hinreichende Eigenschaften von \mathcal{L} -Theorien zu identifizieren, so daß eine formale Bedeutung garantiert werden kann. Dies ist schlußendlich nichts anderes als das Auffinden von Fragmenten des Ausgangsformalismus.

Betrachten wir dafür die klassische Aussagenlogik \mathcal{P} über einem endlichen Alphabet $\{A_1, \dots, A_n\}$ mit der üblichen wahrheitsfunktionalen Semantik $\sigma_{\mathcal{P}}$. Eine gegebene aussagenlogische Theorie muß weder erfüllbar, noch (selbst im Erfüllbarkeitsfalle) ein eindeutig bestimmtes Modell besitzen. Beispiele hierfür sind die einelementigen Theorien $T_1 = \{A_1 \wedge \neg A_1\}$ und $T_2 = \{A_1 \rightarrow A_2\}$ da $\sigma_{\mathcal{P}}(T_1) = \emptyset$ und $|\sigma_{\mathcal{P}}(T_2)| \geq 2$. Die Existenz von Modellen ist garantiert, sofern man zum Beispiel nur noch Mengen bestehend aus Formeln, welche mit Hilfe der Junktoren \vee , \wedge und \rightarrow aufgebaut sind, zuläßt. Hingegen ist die Existenz von genau einem Modell nur dann garantiert sofern man die aussagenlogischen Formeln auf die Atome der assoziierten *Lindenbaum-Tarski Algebra* einschränkt.¹¹ Diese korrespondieren zu genau den n -elementigen Theorien, die jede atomare Variable entweder so wie sie ist oder negiert enthalten. Vergleichbares beobachtet man auch für andere Formalismen. Logikprogramme betrachtet unter der stabilen Semantik können keine, genau eins oder auch mehrere Modelle besitzen. Das Fragment der definiten Logikprogramme garantiert die Existenz als auch die Eindeutigkeit von Modellen.¹²

In dieser Arbeit wurde gezeigt, daß sich abstrakte Argumentationssemantiken ganz ähnlich verhalten. Das Vorhandensein beziehungsweise sogar die Eindeutigkeit von Erweiterungen/Kennzeichnungen hängen im hohen Maße von den syntaktischen Eigenschaften der betrachteten Argumentationsrahmenwerke ab. Um zu zeigen, daß eine konkrete Argumentationssemantik σ für eine bestimmte Menge \mathcal{C} von Rahmenwerken nicht universell definiert ist, das heißt nicht immer eine nicht-leere Semantik zuweist, reicht es aus, einen einzelnen Zeugen $F \in \mathcal{C}$ mit $\sigma(F) = \emptyset$ zu repräsentieren. In diesem Falle spricht man auch vom Kollabieren der Semantik σ .¹³ Für nachfolgendes abstraktes Argumentationsrahmenwerk G kollabiert die stabile Semantik.



Eine bejahende Antwort hingegen, also der Nachweis daß eine Semantik für alle infrage kommenden Rahmenwerke Erweiterungen/Kennzeichnungen liefert ist für gewöhnlich ungleich schwieriger. Der Grund dafür ist, daß man für die gängigen strukturellen Klassen von Rahmenwerken eine Allquantifikation über eine unendliche Menge zu beweisen hat. Im Falle von unbeschränkten Rahmenwerken wird zum

¹¹ Alfred Tarski und John Corcoran, 1983, *Logic, Semantics, Metamathematics: Papers From 1923 to 1938*, Hackett Pub. Co.

¹² Gabbay et al., 1998, *Handbook of Logic in Artificial Intelligence and Logic Programming: Volume 5: Logic Programming*, Oxford University Press.

¹³ Christof Spanring, 2015, *Hunt for the Collapse of Semantics in Infinite Abstract Argumentation Frameworks*, ICCSW, 70–77.

Nachweis der Existenz oft das zum *Auswahlaxiom* äquivalente *Lemma von Zorn*¹⁴ bemüht, wohingegen für endliche Rahmenwerke ein Nachweis oftmals noch "zu Fuß" gelingt, da unendlich aufsteigende Teilmengenkettens schlicht und ergreifend nicht existieren können. Aus der Gesamtheit der Ergebnisse sollen noch zwei durchaus unerwartete erwähnt werden: Erstens, die sogenannte *eifrige Semantik*¹⁵ nimmt eine herausragende Stellung unter den betrachteten Semantiken ein. Entweder gibt es genau eine oder aber unendlich viele eifrige Erweiterungen/Kennzeichnungen. Zweitens, die sogenannte *arrangierte Semantik*¹⁶ garantiert die Existenz von Erweiterungen/Kennzeichnungen nicht nur für endliche Rahmenwerke, sondern auch für jene unendliche, bei denen jedes Argument von maximal nur endliche vielen anderen attackiert wird.

Ausdruckskraft

Wenn ein logischer Formalismus \mathcal{L} zur Wissensrepräsentation beziehungsweise als Modellierungswerkzeug im Allgemeinen eingesetzt wird, besteht ein großes Interesse daran, zu wissen welche Mengen von Modellen überhaupt darstellbar sind. Formeller, sofern $\sigma_{\mathcal{L}}$ die Semantik von \mathcal{L} beschreibt, möchten wir Kenntnis über die Menge $\mathcal{R}_{\mathcal{L}} = \{\sigma_{\mathcal{L}}(T) \mid T \text{ ist eine } \mathcal{L}\text{-theory}\}$ erlangen. Dieses Problem ist auch unter dem Namen *Realisierbarkeit* beziehungsweise *Definierbarkeit* bekannt. Mögliche notwendige oder hinreichende Eigenschaften für das Enthaltensein in $\mathcal{R}_{\mathcal{L}}$, das heißt $\sigma_{\mathcal{L}}$ -realisierbar zu sein, können eine Logik sofort unbrauchbar oder perfekt passend für eine bestimmte Anwendung erscheinen lassen.

Im Falle der klassischen Aussagenlogik \mathcal{P} über einem endlichen Alphabet kann bedingungslos jede Menge von zweiwertigen Belegungen realisiert werden. Das heißt, für eine beliebige Menge von Belegungen \mathcal{I} finden wir immer eine Menge von Formeln T mit der Eigenschaft $\sigma_{\mathcal{P}}(T) = \mathcal{I}$. Dies ist nicht der Fall für normale Logikprogramme betrachtet unter der stabilen Modellsemantik. Hier gilt, daß eine endliche Kandidatenmenge realisierbar ist, sofern diese eine Teilmengenantikette bildet. Dies bedeutet, daß je zwei Mengen aus der Kandidatenmenge unvergleichbar bezüglich der Teilmengenrelation sein müssen. Bemerkenswert ist, daß die Antiketteneigenschaft nicht nur als notwendige, sondern sogar als hinreichende Eigenschaft für Realisierbarkeit auftritt.¹⁷ Ein typisches Gebiet in welchem Realisierbarkeitswissen nützlich ist, ist das Feld der *Wissensrevision*.¹⁸ Grob gesagt bearbeitet dieser Forschungszweig das Problem der Integration von neuer Information in eine Wissensbasis. Die initiale Wissensbasis wird hierbei durch eine \mathcal{L} -Theorie T repräsentiert. Die Zielwissensbasis wird für gewöhnlich semantisch spezifiziert, das heißt gesucht wird eine \mathcal{L} -Theorie S die für eine, durch einen konkreten Revisionsoperator festgelegte, Modellmenge M die Gleichung $\sigma_{\mathcal{L}}(S) = M$ erfüllt. Vor dem Versuch, die Revision auf eine gewisse minimale Art und Weise durchzuführen, ist es vernünftig zu prüfen, ob M überhaupt $\sigma_{\mathcal{L}}$ -realisierbar ist. Anders ausgedrückt: Gilt $M \in \mathcal{R}_{\mathcal{L}}$?

Die ersten Realisierbarkeitsuntersuchungen für Argumentationssemantiken begannen vor einer halben Dekade. Paul E. Dunne und seine Wiener Kollegen prägten den Begriff der *Signatur* für die Menge aller realisierbaren Erweiterungen.¹⁹ Um zu zeigen, daß eine bestimmte Menge von Erweiterungen zu einer Signatur gehört, reicht es aus, ein bezeugendes abstraktes Argumentationsrahmenwerk zu präsentieren. Zum Beispiel, bestätigen die Rahmenwerke F und G , daß die Mengen $\{a_1, b_2, b_3\}, \{a_2, b_1, b_3\}$,

¹⁴Max Zorn, 1935, *A remark on method in transfinite algebra*, American Mathematical Society, 41:667–670.

¹⁵Martin Caminada, 2007, *Comparing Two Unique Extension Semantics for Formal Argumentation: Ideal and Eager*, BNAIC, 81–87.

¹⁶Bart Verheij, 1996, *Two Approaches to Dialectical Argumentation: Admissible Sets and Argumentation Stages*, FAPR, 357–368.

¹⁷Eiter et al., 2013, *Model-based recasting in answer-set programming*, Journal of Applied Non-Classical Logics, 23(1–2):75–104.

¹⁸Alchourrón et al., 1985, *On the logic of theory change: Partial meet contraction and revision functions*, Journal of Symbolic Logic, 50:510–530.

¹⁹Dunne et al., 2013, *Characteristics of multiple viewpoints in abstract argumentation*, Workshop on Dynamics of Knowledge and Belief, 16–30.

$\{a_3, b_1, b_2\}$ als auch $\{\emptyset\}$ Elemente der stabilen Signatur sind. Um zu zeigen, daß dieses Enthalten-sein auf eine Menge, wie zum Beispiel $\{\{a, b\}, \{a, c\}, \{b, c\}\}$, nicht zutrifft, muß man eine Aussage über unendlich viele Rahmenwerke beweisen. Genau dafür stellten die Kollegen lokale Charakterisierungseigenschaften bereit. Diese sind notwendige als auch hinreichende Kriterien bezüglich des Enthaltenseins in einer Signatur, welche sich ausschließlich auf Eigenschaften der Kandidatenmenge beziehen.

In dieser Arbeit präsentieren wir Charakterisierungseigenschaften für eine repräsentative Anzahl von Argumentationssemantiken. Die Übersicht beinhaltet auch semantisch eingeschränkten Versionen von Realisierbarkeit, nämlich die der *kompakten* und die der *analytische Realisierbarkeit*.²⁰ Beide Versionen sind durch typisch auftretende Phänomene, beobachtbar bei verschiedenen Semantiken, motiviert. Erstens, das Auftreten von *global widerlegten Argumenten*. Dies sind Argumente eines Rahmenwerkes, welche in keiner einzigen Erweiterung enthalten sind. Zweitens, die Möglichkeit von *impliziten Konflikten* unter einer Argumentationssemantik. Ein impliziter Konflikt zwischen zwei Argumenten liegt vor, wenn diese niemals zusammen in einer Erweiterung vorkommen, obwohl kein Argument das andere attackiert. Um zu verstehen, auf welche Art und Weise global widerlegte Argumente als auch implizite Konflikte zur Ausdruckskraft einer Semantik beitragen, wurden die sogenannten *kompakten* und des Weiteren, die *analytischen Argumentationsrahmenwerke* eingeführt. Die erste Art verbietet das Vorkommen von global widerlegten Argumenten, während die letztere frei von impliziten Konflikten ist. Es stellt sich heraus, daß für eine Vielzahl der betrachteten Semantiken die volle Ausdruckskraft tatsächlich auf der Verwendung global widerlegter Argumente und impliziter Konflikte beruht. Dies bedeutet leider auch, daß diese gut greifbaren Klassen von Rahmenwerken nicht zur Definition einer Normalform verwendet werden können. Eine erste Studie der Ausdruckskraft für uneingeschränkte Rahmenwerke zeigt einen engen Zusammenhang zu der Ausdruckskraft endlicher und dabei kompakter Rahmenwerke auf. Eine tiefgreifende Analyse steht jedoch noch aus. Zum Beispiel konnten noch keine vollständigen Charakterisierungseigenschaften für die Realisierbarkeit im Unendlichen identifiziert werden.

Ersetzbarkeit

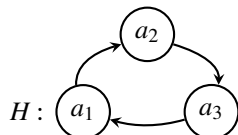
Gegeben zwei syntaktisch verschiedene \mathcal{L} -Theorien T_1 und T_2 . Inwiefern ist es möglich, daß diese beiden \mathcal{L} -Theorien dieselbe Information darstellen? Um diese Frage zu beantworten, müssen wir vorab klarstellen, was wir genau meinen mit, *dieselbe Information zu tragen*. Zuerst ist es wichtig festzustellen, daß es weder eine eindeutig bestimmte Antwort, noch irgendeine vom Formalismus \mathcal{L} bevorzugte Antwort darauf gibt. Zum Beispiel, führt das Gleichsetzen von Informationen einer Theorie mit deren Semantik zum weitverbreiteten Begriff der *gewöhnlichen Äquivalenz*. Präziser, zwei \mathcal{L} -Theorien T_1 und T_2 sind dann und nur dann äquivalent (besitzen also dieselbe Information), wenn $\sigma_{\mathcal{L}}(T_1) = \sigma_{\mathcal{L}}(T_2)$. Eine schärfere Variante des Tragens der selben Informationen ist die semantische Ununterscheidbarkeit von beiden Theorien bezüglich der simultanen Erweiterung durch beliebige andere Theorien. Formeller ausgedrückt können wir sagen: T_1 und T_2 werden als äquivalent angesehen, dann und nur dann, wenn $\sigma_{\mathcal{L}}(T_1 \cup T) = \sigma_{\mathcal{L}}(T_2 \cup T)$ für jede \mathcal{L} -Theorie T . Diese Interpretation vom Tragen der gleichen Information wird als *starke Äquivalenz* bezeichnet und ist für jeden logischen Formalismus von großem Interesse. Der Grund dafür liegt darin, daß starke Äquivalenz es ermöglicht, Teile einer gegebenen Theorie lokal zu ersetzen und möglicherweise dabei zu vereinfachen, ohne dabei die Semantik der Ausgangstheorie zu verändern.

Für die klassische Aussagenlogik und die Prädikatenlogik 1. Stufe fallen gewöhnliche und starke Äquivalenz zusammen. Dies ist bemerkenswert, da rein definitorisch nur das implizieren von gewöhnlicher, bei gegebener starker Äquivalenz gerechtfertigt ist. An dieser Stelle sei erwähnt, daß es nicht die

²⁰Baumann et al. 2016 *On rejected arguments and implicit conflicts: The hidden power of argumentation semantics*, Artificial Intelligence, 241:244–284.

Monotonie ist, welche das Zusammenfallen beider Äquivalenzbegriffe verursacht.²¹ Für alle in der Literatur verfügbaren nichtmonotonen Formalismen kann man Theorien beziehungsweise Objekte finden, welche zwar gewöhnlich, aber nicht stark äquivalent sind. Aufgrunddessen wurden große Anstrengungen unternommen, die starke Äquivalenz für nichtmonotone Formalismen und Logiken im Allgemeinen zu charakterisieren.²²

Vor einer knappen Dekade begann die Forschung zur starken Äquivalenz für abstrakte Argumentationsrahmenwerke. Es ist nicht schwer zu überprüfen, daß nachfolgendes Rahmenwerk H keine stabilen Erweiterungen besitzt. Demzufolge gilt, daß die Rahmenwerke G und H gewöhnlich äquivalent bezüglich der stabilen Semantik sind.



Um nun zu zeigen, daß beide Rahmenwerke jedoch nicht stark äquivalent sind, besteht ohne vertiefte Forschung der Thematik nur die Möglichkeit der Suche nach einem bezeugenden Rahmenwerk Z . Betrachten wir dazu nachfolgende Vereinigungen (Vereinigungen von Rahmenwerken werden punktweise verstanden).



Das Rahmenwerk $H \cup Z$ besitzt nur die Menge $\{a_3, a_4\}$ als stabile Erweiterung, wobei hingegen für $G \cup Z$ nur die Menge $\{a_1, a_4\}$ die Anforderungen der stabilen Semantik erfüllt. Es ist das erstaunliche Resultat von Oikarinen und Woltran, welches uns auf eine aufwendige Suche von Zeugen verzichten läßt.²³ Es wurde nämlich gezeigt, daß obwohl starke Äquivalenz rein semantisch definiert ist, ein rein syntaktischer Vergleich zur Beantwortung der Frage ausreicht. Dafür wurde das Konzept des *Kerns* eines Rahmenwerks, welcher ein Teilgraph des Ausgangsrahmenwerkes ist, eingeführt. Informell gesprochen, stellt ein Kern eine redundanzfreie Version des initialen Rahmenwerks dar. Es konnte gezeigt werden, daß die syntaktische Identität solcher Kerne notwendig und hinreichend für die starke Äquivalenz von Rahmenwerken ist. Im weiteren Verlauf der Forschung wurde festgestellt, daß in Bezug auf dynamische Entwicklungen von Rahmenwerken, die starke Äquivalenz, eine unnötig starke Anforderung ist, da für viele Anwendungen die Art der Veränderung von vornherein bestimmt ist, und keinesfalls alle denkbaren Szenarien ausgeschöpft werden.²⁴

In dieser Arbeit werden verschiedenste Äquivalenzbegriffe definiert und charakterisiert. Ein Vergleich der Resultate zwischen erweiterungsbasierten und kennzeichnungsbasierten Semantiken ergibt,

²¹Ringo Baumann and Hannes Straß, 2016, *An Abstract Logical Approach to Characterizing Strong Equivalence in Logic-based Knowledge Representation Formalisms*, Principles of Knowledge Representation and Reasoning, 525–528.

²²Mirosław Truszczyński, 2006, *Strong and uniform equivalence of nonmonotonic theories - an algebraic approach*, Annals of Mathematics and Artificial Intelligence, 48:245–265.

²³Emilia Oikarinen and Stefan Woltran, 2011, *Characterizing strong equivalence for argumentation frameworks*, Artificial Intelligence, 175:1985–2009.

²⁴Ringo Baumann, 2012, *Normal and strong expansion equivalence for argumentation frameworks*, Artificial Intelligence, 193:18–44.

das letztere sich uniformer bezüglich charakterisierender Kerne verhalten. Im Gegensatz zu erweiterungsbasierten Semantiken, welche verschiedene, als auch stellenweise sehr aufwendig beschriebene Kerne verwenden, benötigen kennzeichnungs-basierte Semantiken nur eine geringe Anzahl an einfach zugänglichen Kernen. Interessant ist auch, daß es für einige Semantiken keine Unterschiede zwischen der Charakterisierung im endlichen und unbeschränkten Fall gibt. Für Semantiken, die dies im Allgemeinen nicht gewährleisten, können hinreichende Bedingungen für das Zusammenfallen angegeben werden. Eine von diesen Bedingungen ist die Forderung nach der *gemeinsamen Erweiterbarkeit*.²⁵

Überprüfbarkeit

Überprüfbarkeit für Rahmenwerke, in dem Sinne wie es hier studiert wurde, hat kein offensichtliches Gegenstück in anderen bekannten logischen Formalismen. Abstrakte Rahmenwerke sind mathematisch gesehen gerichtete Graphen und Argumentationssemantiken bilden eben diese Graphen auf Mengen von Mengen von Argumenten ab. Stellen wir uns vor, daß eine künstliche Intelligenz ein abstraktes Rahmenwerk zur Wissensrepräsentation benutzt und ihr Handeln einzig und allein von den Erweiterungen dieses Rahmenwerks abhängt. Wenn wir uns nun weiter vorstellen, daß unser zukünftiges Handeln im hohem Maße von der Kenntnis der Erweiterungen des uns nicht zugänglichen Rahmenwerks abhängt, drängt sich folgende Frage sofort auf: Kann man eine Argumentationssemantik σ nur dann berechnen, wenn das gesamte Rahmenwerk zur Verfügung steht? Oder anders ausgedrückt: Ist es möglich, akzeptable Mengen bezüglich σ eindeutig zu berechnen, auch wenn man echt weniger Information als das gesamte Rahmenwerk zur Verfügung hat? Folgende Überlegungen sind zielführend.

1. Eine Grundvoraussetzung für fast alle existierenden Semantiken ist die Konfliktfreiheit, das heißt, Argumente innerhalb einer Erweiterung dürfen sich nicht angreifen. Konsequenterweise, sollte das Wissen über konfliktfreie Mengen ein wesentlicher Bestandteil zur Berechnung einer Semantik sein.
2. Der zweite Schritt ist die Beantwortung der Frage: Was für Informationen benötigt man zusätzlich zu konfliktfreien Mengen? Gegeben folgende Menge $M = \{\emptyset, \{a\}, \{b\}\}$. Sofern M die gesamten konfliktfreien Mengen darstellt, können wir schließen, daß zwischen a und b mindestens eine Attacke vorliegen muß. Die folgenden drei Rahmenwerke besitzen genau die in M spezifizierten Mengen als konfliktfreie, werten sich jedoch unterschiedlich in Bezug auf gängige Semantiken aus.



Zum Beispiel sind $\{a\}$ als auch $\{b\}$ stabile Erweiterungen von C_3 , wohingegen nur jeweils eine von ihnen akzeptabel in C_1 beziehungsweise C_2 ist. Demzufolge scheint Wissen über die Umgebung von Argumenten unabdingbar zu sein.

3. Der letzte Schritt ist einer der Minimierung. Mit anderen Worten, welche Art von Wissen über die Nachbarschaft ist angesichts der Berechnung entbehrlich? Dies hängt natürlich stark von der betrachteten Semantik ab. Zum Beispiel stellt sich heraus, daß wir für die stabile Semantik keine Informationen über eingehende Angriffe eines Arguments benötigen.

²⁵Ringo Baumann and Chirstof Spanring, 2017, *A study of unrestricted abstract argumentation frameworks*, International Joint Conference on Artificial Intelligence, 807–813.

Ein Hauptresultat dieser Arbeit ist, daß alle betrachteten Semantiken tatsächlich mit weniger Informationen als das gesamte Rahmenwerk berechnet werden können. Die Menge der benötigten Information wird dabei in sogenannte *Überprüfbarkeitsklassen* unterschieden. Diese können durch eine Informationsordnung in Bezug zu einander gesetzt werden. Interessanterweise fallen viele syntaktisch verschiedene Klassen bezüglich der Ordnung zusammen. Des Weiteren wird von diesen wenigen semantisch unterschiedlichen Überprüfbarkeitsklassen wiederum nur ein Bruchteil zur Überprüfbarkeit aller gängigen Semantiken gebraucht. Als Nebeneffekt der genauen Spezifizierung der Überprüfbarkeit konnten wir frühere Charakterisierungssätze bezüglich starker Äquivalenz verallgemeinern und zum ersten Mal Resultate für sogenannte *intermediäre Semantiken*²⁶ zeigen.

Charakterisierungslogiken für Logische Formalismen

Der zweite Hauptteil dieser Habilitationsschrift beschäftigt sich mit der Existenz und Beschaffenheit von Charakterisierungslogiken für beliebige logische Formalismen \mathcal{L} .

Wie schon vorher erwähnt, gelang es in einer Reihe interessanter Studien, die starke Äquivalenz für mehrere nicht-monotone Formalismen genau zu charakterisieren. Diese Ergebnisse wurden zu Recht als wichtig für das Studium dieser konkreten Wissensrepräsentationsformalismen angesehen, da die Charakterisierung starker Äquivalenz ein tieferes Verständnis der kontextabhängigen Bedeutung von \mathcal{L} -Theorien ermöglicht. Ein bemerkenswertes Resultat darunter ist die Charakterisierung für normale Logikprogramme unter der stabilen Modellsemantik. Es konnte gezeigt werden, daß zwei normale Logikprogramme unter stabilen Modellen genau dann stark äquivalent sind, sofern eben jene Programme in der monotonen Logik *Hier und Dort* gewöhnlich äquivalent sind.²⁷ Dieses Charakterisierungsergebnis wirft sofort folgende grundlegende Frage auf:

Besitzt jeder Formalismus \mathcal{L} eine klassische Charakterisierungslogik \mathcal{L}' ?

Etwas formaler ist dies die Frage nach einer Logik \mathcal{L}' , sodaß für alle \mathcal{L} -Theorien T_1, T_2 gilt: $\sigma(T_1 \cup T) = \sigma(T_2 \cup T)$ für jede \mathcal{L} -Theorie T genau dann, wenn $\sigma'(T_1) = \sigma'(T_2)$. Bevor wir genauer ausführen, was wir unter einer klassischen Logik verstehen wollen (solange appellieren wir an des Lesers Intuition), gehen wir auf eben gestellte Frage ein.

Es ist das zentrale Ergebnis dieser Habilitationsschrift, daß die Antwort auf diese Frage ein zwar bedingtes, aber wohlwollendes "Ja" ist. Genauer gesagt, obgleich nicht jeder Formalismus eine Charakterisierungslogik besitzt, konnten wir zeigen, daß der wichtige Fall der Einschränkung auf endliche Wissensbasen die Existenz garantiert, und dies, in einem sehr allgemeinen Kontext. Frühere Ergebnisse bei der Charakterisierung von Formalismen nutzen die Besonderheiten eines jeden Formalismus aus. Unser vorgestellter Ansatz abstrahiert vollständig von der Art der syntaktischen Bausteine des Formalismus sowie der konkreten Ausgestaltung der Modelltheorie. Unsere Resultate zeigen nicht nur die Existenz irgendeines charakterisierenden Formalismus, sondern vielmehr, daß die Modelltheorien möglicher Charakterisierungslogiken für einen konkreten Formalismus eindeutig modulo Isomorphie bestimmt sind und darüber hinaus, strukturell einer klassischen Logik ähneln.

Unter einer *klassischen Logik* verstehen wir genau diejenigen Logiken, deren Semantiken die *Schnitteigenschaft* erfüllen, das heißt $\sigma'(T) = \bigcap_{t \in T} \sigma'(\{t\})$ für jede \mathcal{L}' -theorie T . Diese Eigenschaft garantiert die Monotonie der assoziierten Konsequenzrelation und ist des Weiteren äquivalent dazu, daß es eine Theoriefunktion τ gibt, für welche σ und τ in *Galoiskorrespondenz*²⁸ stehen. Viele gut untersuchte

²⁶Pietro Baroni and Massimiliano Giacomin, 2007, *Comparing argumentation semantics with respect to skepticism*, Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Computer Science, 210–221

²⁷Lifschitz et al., 2001, *Strongly equivalent logic programs*, ACM Transactions on Computational Logic, 2(4):526–541.

²⁸Brain A. Davey und Hilary A. Priestley, 2002, *Introduction to Lattices and Order (Second edition)*, Cambridge University Press.

Logiken (die wir allein schon aufgrund ihrer Allgegenwart klassisch nennen würden) besitzen die Schnitteigenschaft einfach per Definition, da ihre semantischen Grundbausteine Formeln statt Theorien sind.

Abschließend möchten wir darauf hinweisen, daß der Nachweis der Existenz von Charakterisierungslogiken durch eine kanonische Konstruktion, welche stark an *Herbrand-modelle*²⁹ erinnert, gelang. Genauer gesagt, wird die Charakterisierungssemantik durch Vereinigungen von bestimmten starken Äquivalenzklassen des Ausgangsformalismus definiert. Dies impliziert auch, daß die kanonische Modellsemantik weder intuitiv noch selbsterklärend ist. Die beabsichtigte Rolle dieser Semantik ist einzig und allein die eines Zeugens. Nichtsdestotrotz garantieren unsere Resultate, daß die Suche nach intuitiven Ziellogiken für konkrete Wissensrepräsentationsformalismen, wie im Falle der Logik von Hier und Dort bezüglich Logikprogrammen unter der stabilen Modellsemantik, nicht zum Scheitern verurteilt ist.

Ausblick

Die öffentliche und wissenschaftliche Meinung darüber, zu was die Forschung zur Künstliche Intelligenz in der Lage ist zu erreichen, hat sich seit dem mittlerweile berühmten *Dartmouth Sommer Seminar*³⁰, welches von den meisten Forschern als Beginn der KI-Forschung betrachtet wird, häufig geändert. Die Geschichte der KI, ist eine Geschichte voller Höhen und Tiefen. Phasen der Begeisterung, die mit breiten Finanzierungsmöglichkeiten ausgestattet waren und Etappen der Enttäuschung und Skepsis, sogenannte *KI Winter*, wechselten sich stetig ab. Momentan befinden wir uns definitiv in einer Hochphase. Mehrere führende KI Konferenzen haben beschlossen, statt zweijährlich, von nun an jährlich stattzufinden, darunter auch die renommierte *Internationale Konferenz für künstliche Intelligenz*³¹ (ab 2015), sowie die *Internationale Konferenz für Prinzipien der Wissensrepräsentation und des Schließens*³² (ab 2020). Darüber hinaus hat die Deutsche Bundesregierung beschlossen, bis zum Jahre 2025 mehr als 3 Milliarden Euro für die Forschung und Entwicklung der künstlichen Intelligenz auszugeben. Um die Fachkräfteabwanderung zu stoppen, wird diese Initiative eine Finanzierung von circa 100 neuen Professuren umfassen. Hiermit ist das Ziel klar formuliert: Deutschland will bei der KI weltweit führend werden.

Ich bin mir sicher, daß die nächsten Jahre eine spannende Zeit für die Künstliche Intelligenz sein werden. Eine der größten Herausforderungen wird darin bestehen, Forschungsergebnisse in die Industrie und letztendlich in unser tägliches Leben sinnvoll zu transferieren. Es ist nicht zu weit hergedacht, sich vorzustellen, daß die Expertise von Programmen wie AlphaZero und Watson dazu verwendet werden könnten, um zum Beispiel gefährliche Missionen wie Brandbekämpfung oder auch die Unterstützung von Menschen bei täglichen Entscheidungen zu optimieren beziehungsweise zu unterstützen. Was bisher von beiden Programmen nicht erreicht wurde, ist die Fähigkeit zu erklären, warum bestimmte Entscheidungen getroffen wurden. Dies bedeutet, daß die von diesen Künstlichen Intelligenzen durchgeführten Aktionen weder transparent noch für den Menschen verständlich sind. Um eine breite Zustimmung für KI-Technologien zu erhalten, müssen Erklärungskomponenten implementiert werden. Dieses Problem beziehungsweise diese Anforderung wird als *erklärbare KI* bezeichnet.³³ Die Argumentationstheorie liefert die formalen Voraussetzungen um eine Diskussion zu führen oder einen bestimmten Vorschlag zu rechtfertigen, da Entscheidungen aufgrund von Pro- und Kontra-Argumenten gefällt werden. Aufgrunddessen glaube ich, daß vor allem die vorgestellte formale Analyse der intrinsischen Eigenschaften der abstrakten Argumentationsrahmenwerke zur Entwicklung einer neuen KI-Generation beitragen kann.

²⁹Ebbinghaus et al., 1996, *Einführung in die mathematische Logik*, Spektrum Akademischer Verlag

³⁰McCarthy et al., 1955, *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>

³¹*International Joint Conferences on Artificial Intelligence Organization*, 2019, <https://www.ijcai.org/>

³²*Principles of Knowledge Representation and Reasoning*, 2019, <http://www.kr.org/>

³³Andreas Holzinger, 2018, *Explainable AI (ex-AI)*, <https://gi.de/informatiklexikon/explainable-ai-ex-ai/>

Contents

1	Introduction	1
1.1	Publications	3
2	On the Existence of Characterization Logics for Knowledge Representation Formalisms	5
2.1	An Abstract View on Model Theory	6
2.1.1	Models and Consequences	7
2.1.2	Standard and strong equivalence	10
2.1.3	Galois correspondences	13
2.2	Characterization Logics	17
2.2.1	Covered Logics	22
2.2.2	Finite-Theory Characterization Logics	27
2.3	Applying Canonical Constructions to Nonmonotonic Formalisms	31
2.3.1	Abstract Argumentation Theory	31
2.3.2	Normal Logic Programs	37
2.4	Discussion	38
3	Existence and Uniqueness of Argumentation Semantics	39
3.1	Abstract Argumentation Theory	39
3.1.1	Important Syntactical Classes of AFs	40
3.1.2	Argumentation Semantics - The Flagship of Dung's Abstract Argumentation Theory	40
3.1.3	Definedness Statuses of Argumentation Semantics	43
3.2	Existence and Uniqueness w.r.t. Finite AFs	44
3.3	Existence and Uniqueness w.r.t. Arbitrary AFs	45
3.3.1	Non-well-defined Semantics	45
3.3.2	Collapsing Semantics	46
3.3.3	Universally Defined Semantics	48
3.3.4	Uniquely Defined Semantics	50
3.4	Existence and Uniqueness w.r.t. Finitary AFs	50
3.5	Summary of Results and Conclusion	51
4	Expressibility of Argumentation Semantics	53
4.1	Realizability and Signatures	54
4.2	Signatures w.r.t. Finite AFs	55
4.2.1	Semantics based on Conflict-freeness	56
4.2.2	Semantics based on Admissibility	60
4.2.3	Uniquely Defined Semantics	64
4.2.4	Summary of Results and Further Remarks	64
4.3	Signatures w.r.t. Finite, Compact AFs	66

4.3.1	Central Definitions and Preliminary Observations	67
4.3.2	The Loss or Stability of Expressive Power	69
4.3.3	Comparing Finite, Compact Signatures and Final Remarks	70
4.4	Signatures w.r.t. Finite, Analytic AFs	71
4.4.1	Central Definitions and Preliminary Observations	72
4.4.2	The Loss or Stability of Expressive Power	74
4.4.3	Comparing Finite, Analytic Signatures and Final Remarks	76
4.5	Remarks on Unrestricted AFs and Intertranslatability	78
4.6	Realizability and Signatures for Labelling-based Versions	80
4.6.1	Preliminary Results for Labelling-based Signatures	81
4.6.2	Realizability under Projection	82
4.7	Final Remarks and Conclusion	83
5	Replaceability of Subframeworks	85
5.1	Dynamic Scenarios and Corresponding Equivalence Notions	86
5.2	Characterization Theorems for Extension-based Semantics	90
5.2.1	The Central Notion of Expansion Equivalence	90
5.2.2	Further Equivalence Notions Characterizable through Kernels	95
5.2.3	The Exceptional Case of Normal Deletion Equivalence	100
5.2.4	Characterization Theorems in Case of Self-loop-free AFs	102
5.2.5	Summary of Results and Conclusion	102
5.3	Equivalence in the Light of Unrestricted Frameworks	105
5.4	Characterization Theorems for Labelling-Based Semantics	106
5.4.1	Basic Properties and a Fundamental Relation	106
5.4.2	Coincidences of Extension-based and Labelling-based Versions	108
5.4.3	Non-Coincidence of Extension-based and Labelling-based Versions	108
5.4.4	Summary of Results and Conclusion	111
5.5	Final Remarks	112
6	Verifiability of Argumentation Semantics	113
6.1	Preliminaries and Strongly Admissible Sets	114
6.2	Verifiability	117
6.3	Intermediate Semantics	123
6.4	Conclusions	125
7	Summary and Final Remarks	127

List of Figures

3.1	Subset Relations between Semantics	42
4.1	Subset Relations between Finite Signatures	65
4.2	Subset Relations between Finite, Compact AFs	68
4.3	Subset Relations between Finite, Compact Signatures	70
4.4	Subset Relations between Finite, Analytic AFs	73
4.5	Subset Relations between Finite, Analytic Signatures	77
4.6	Translational (Im)Possibilities	80
4.7	Subset Relations between Different Kinds of Signatures	84
5.1	Different Kinds of Expansions	87
5.2	An Update and Different Kinds of Deletions	88
5.3	Preliminary Subset Relations between Equivalence Notions	91
5.4	Extension-based Characterizations for Finite AFs	103
5.5	Relations for $\sigma \in \{stg, stb, ss, eg, ad, pr, il, gr, co, na, cf2, stg2\}$ - Extension-based Versions and Finite AFs	104
5.6	Stable Semantics - Extension-based Version and Finite AFs	104
5.7	Labelling-based Characterizations for Finite AFs	111
5.8	Relations for $\sigma \in \{stb, ss, eg, ad, pr, il, gr, co\}$ - Labelling-based Versions and Finite AFs	112
6.1	Representatives of neighborhood functions and their relation w.r.t. information; a node x stands for the neighborhood function τ^x ; an arrow from x to y means $\tau^x \prec \tau^y$	118
6.2	Semantics and their Exact Verification Classes	122

Chapter 1

Introduction

Given the large variety of existing logical formalisms it is of utmost importance to select the most adequate one for a specific purpose, e.g. for representing the knowledge relevant for a particular application or for using the formalism as a modeling tool for problem solving. Awareness of the nature of a logical formalism or, in other words, of its fundamental intrinsic properties, is indispensable and provides the basis of an informed choice. Apart from a deeper understanding of the considered formalism, the study of such intrinsic properties can help to identify interesting fragments or to develop useful extensions of a formalism. Moreover, the obtained insights can be used to refine existing algorithms or even give rise to new ones.

Presumably, the best-known intrinsic property of logics is *monotonicity*. *Monotonic logics* like propositional logic or first order logic are perfectly suitable for the formalization of universal truths since in these logics, whenever a formula ϕ is a logical consequence of a theory T , it remains true forever and without any exception even if we add further information to T . In contrast, formalisms which do not satisfy monotonicity, commonly referred to as *nonmonotonic logics*, allow for defeasible reasoning, i.e. it is possible to withdraw former conclusions (cf. [Brewka, 1989; Gabbay et al., 1994] for excellent overviews). Both kinds of logics have their traditional application domains and apart from this fundamental choice there are many other criteria of comparison influencing the decision which logic or which specific semantics of a logic to use in a certain context.

One of the first intrinsic properties which comes to mind is *computational complexity*, i.e. how expensive it is to solve typical decision problems in the candidate formalism. A related issue is *modularity* which is, among other things, concerned with the question of whether it is possible to divide a given theory in subtheories, s.t. the formal semantics of the entire theory can be obtained by constructing the semantics of the subtheories. Both topics were studied in-depth for mainstream nonmonotonic formalisms such as default logic [Gottlob, 1992; Turner, 1996], logic programming under certain semantics [Dantsin et al., 1997; Lifschitz and Turner, 1994] as well as abstract argumentation frameworks (AFs) under various argumentation semantics [Dvořák and Dunne, 2018; Baroni et al., 2018b].

In this habilitation treatise we elaborate a theory of four further intrinsic properties of abstract argumentation semantics. In brief, Dung-style AFs consist of arguments and attacks which are treated as primitives, i.e., the internal structure of arguments is not considered. The major focus is on resolving conflicts. To this end a variety of semantics have been defined, each of them specifying acceptable sets of arguments, so-called extensions, in a particular way. Another main approach to argumentation semantics are so-called labelling-based semantics which contain more information than their extension-based counterparts. A labelling explicitly classifies each argument either as accepted, rejected or undecided.

-
1. *existence and uniqueness* Is it possible, and if so how, to *guarantee* the existence of at least one or exactly one extension/labelling by considering the structure of a given AF F only? (Chapter 3)
 2. *expressibility* Is it possible, and if so how, to *realize* a given candidate set of extensions/labellings within a single AF F ? (Chapter 4)
 3. *replaceability* Is it possible, and if so how, to *simplify* parts of a given AF F s.t. the modified version F' and F cannot be distinguished semantically by further information that might be added later to both simultaneously? (Chapter 5)
 4. *verifiability* Is it possible, and if so how, to *compute* the semantics of an AF F unambiguously, given that we are faced with strictly less information than the entire framework F ? (Chapter 6)

The question whether a certain formalism always provides one with a formal meaning or even with a uniquely determined semantical answer is a crucial factor for its suitability for the application in mind. For instance, in problem solving a plurality of solutions may possibly be desired, whereas in decision making one might be interested in guaranteeing a single answer. It is well-known that a given theory in propositional logic neither has to possess a model nor, in case of existence, has there to be exactly one. The same applies to logic programs under stable model semantics. In contrast, a propositional theory of only positive formulae is always satisfiable and definite logic programs constitute a subclass of logic programs where even uniqueness of a model is guaranteed. In Chapter 3 we will see that Dung’s abstract argumentation semantics behave in a similar way, i.e. the existence and uniqueness of extensions/labellings depend on structural restrictions of argumentation frameworks.

Expressibility is concerned with the expressive power of logical formalisms. The question here is which kinds of models are realizable, that is, can be the set of models of a single knowledge base of the formalism. This is a decisive property from an application angle since potential necessary or sufficient properties of model sets may rule out a logic or make it perfectly appropriate for representing certain solutions. For instance, it is well-known that in case of propositional logic every finite set of two-valued interpretations is realizable. This means, given such a finite set \mathcal{I} , we always find a set of formulae T , s.t. $\text{Mod}(T) = \mathcal{I}$. In case of normal logic programs it is obvious that not all model sets can be expressed, since any set of stable models forms a \subseteq -antichain. Remarkably, being such an antichain is not only necessary but even sufficient for realizability w.r.t. stable model semantics [Eiter et al., 2013; Strass, 2015]. In case of abstract argumentation we are equipped with a high number of semantics, for which we will see in Chapter 4 that characterizing properties are not that easy to find. Moreover, as expected, representational limits highly depend on the chosen semantics.

In case of propositional logic – in contrast to all nonmonotonic logics available in the literature – we have that standard equivalence, i.e. sharing the same models, even guarantees intersubstitutability in any logical context without loss of information. As an aside, it is not the monotonicity of a certain logic but rather the so-called *intersection property* which guarantees this behavior [Baumann and Strass, 2016]. Substitutability is of great importance for dynamically evolving scenarios since it allows one to simplify parts of a theory without looking at the rest. For this reason, much effort has been devoted to characterizing *strong equivalence* for nonmonotonic formalisms, such as logic programs [Lifschitz et al., 2001], causal theories [Turner, 2004], default logic [Turner, 2001] and nonmonotonic logics in general [Truszczyński, 2006; Baumann and Strass, 2016]. In Chapter 5 we will see that characterization theorems in case of abstract argumentation are quite different from those for the aforementioned formalisms in that being strongly equivalent can be decided by looking at the syntax only.

Verifiability is a topic that is very specific for abstract argumentation. Over the last two decades a bunch of argumentation semantics were introduced [Baroni et al., 2018a]. The motivations of these semantics range from the desired treatment of specific examples to fulfilling a number of abstract principles. Mathematically speaking, a semantics takes as input an AF and returns acceptable sets of arguments.

Verifiability deals with the question whether we really need the entire AF to compute a certain argumentation semantics σ ? In other words, is it possible to unambiguously determine acceptable sets w.r.t. σ , given only partial information of the underlying framework F . In Chapter 6 we will see that most of the existing semantics can be computed with indeed less information than the entire framework. We will categorize the amount of information required by taking the conflict-free sets as a basis and distinguishing between different amounts of knowledge about the neighborhood of these sets, so-called *verification classes*. These classes allow us to categorize semantics with respect to the information needed to verify whether a certain set of arguments is acceptable. The study of this topic as well as the properties mentioned before pave the way for a more general view on argumentation semantics, their common features, and their inherent differences.

Let us return to knowledge representation formalisms in general. Two knowledge bases are strongly equivalent if and only if they are mutually interchangeable in arbitrary contexts. This notion is of high interest for any logical formalism since it allows one to locally replace, and thus potentially give rise to simplification, parts of a given theory without changing the semantics of the latter. As already said, since it is possible to find ordinarily but not strongly equivalent objects for any nonmonotonic formalism available in the literature a lot of research has been devoted to characterizing strong equivalence. For example, it turned out that strong equivalence for logic programs under stable models can be characterized by so-called HT-models [Lifschitz et al., 2001]. More precisely, two logic programs are strongly equivalent if and only if they are standard equivalent in the logic of here-and-there. This means the logic of here-and-there can be seen as a characterizing formalism for logic programs under stable model semantics. In Chapter 2 we will study whether the existence of such characterization logics can be guaranteed for every logic. One main result is that every knowledge representation formalism that allows for a notion of strong equivalence on its finite knowledge bases also possesses a canonical characterizing formalism. In particular, we argue that those characterizing formalisms can be seen as classical, monotonic logics which are uniquely determined (up to isomorphism) regarding their model theory.

1.1 Publications

Most of the results presented in this habilitation thesis have already been published as conference papers or journal articles, as a chapter of a handbook as well as in a book in honour of Gerhard Brewka. Moreover, an extended version of [Baumann and Strass, 2016] is currently under review in the Journal of Artificial Intelligence Research (JAIR). In the following we list the involved publications together with the chapters or sections where they are mainly used.

- conference papers

1. [Baumann, 2014a] (Sections 5.2.2 and 5.2.3)
Context-free and context-sensitive kernels: Update and deletion equivalence in abstract argumentation
European Conference on Artificial Intelligence (ECAI)
2. [Baumann et al., 2014a] (Section 4.3)
Compact argumentation frameworks
European Conference on Artificial Intelligence (ECAI)
3. [Baumann and Brewka, 2015] (Section 2.3.1)
AGM meets abstract argumentation: Expansion and revision for Dung frameworks
International Joint Conference on Artificial Intelligence (IJCAI)
4. [Baumann, 2016] (Section 5.4)
Characterizing equivalence notions for labelling-based semantics
International Conference on Principles of Knowledge Representation and Reasoning (KR)

5. [Baumann and Strass, 2016] (Chapters 2)
An abstract logical approach to characterizing strong equivalence in logic-based knowledge representation formalisms
International Conference on Principles of Knowledge Representation and Reasoning (KR)
 6. [Baumann et al., 2016b] (Chapter 6)
Verifiability of argumentation semantics
Computational Models of Argument (COMMA)
 7. [Baumann and Spanring, 2017] (Sections 3.3, 4.5 and 5.3)
A study of unrestricted abstract argumentation frameworks
International Joint Conference on Artificial Intelligence (IJCAI)
- journal articles
8. [Baumann et al., 2016a] (Sections 4.3 and 4.4)
On rejected arguments and implicit conflicts: The hidden power of argumentation semantics
Journal of Artificial Intelligence (AIJ)
- book contribution
9. [Baumann, 2018] (Chapters 3, 4, 5 and 6)
On the nature of argumentation semantics: Existence and uniqueness, expressibility, and replaceability
Handbook of Formal Argumentation
 10. [Baumann and Spanring, 2015] (Chapters 3)
Infinite argumentation frameworks - On the existence and uniqueness of extensions
Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation – Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday

Chapter 2

On the Existence of Characterization Logics for Knowledge Representation Formalisms

Reusability of human-made artifacts is of paramount importance in computer science. To assess the reusability of (parts of) knowledge bases in logic-based knowledge representation languages, we have to know whether pieces of knowledge make certain context-dependent assumptions. In classical propositional logic, for example, there is no such context-dependence: whenever two (sets of) formulas are equivalent in the sense of having the same models, then they are mutually replaceable in arbitrary contexts.

In the field of knowledge representation and reasoning, however, not all commonly used formalisms are like classical logic in this regard. For example, the answer-set-programming paradigm uses the formalism of normal logic programs to encode combinatorial problems such that the answer sets (stable models) of the logic programs correspond to and encode solutions to the encoded problem [Gebser et al., 2012]. Alas, for normal logic programs, having the same stable models does *not* amount to mutual replaceability. For this, a stronger property is needed: it is called *strong equivalence*, and holds for two logic programs if and only if they keep the same stable models even if they are both extended with an arbitrary third logic program. Formally – as a logic program is a set of rules –, extending a logic program can be modeled by ordinary set union. Consequently, the notion of strong equivalence can be defined in a similar way for other knowledge representation formalisms for which set union is an adequate formalization of appending or otherwise combining knowledge bases.

In a series of interesting developments, researchers have succeeded in precisely characterizing strong equivalence for several formalisms, among them normal logic programs under the stable model semantics [Lifschitz et al., 2001; Turner, 2001]. What is more, it turned out that sometimes, to characterize strong equivalence in formalism F , we can use ordinary equivalence in formalism F' : for example, strong equivalence in normal logic programs under stable models can be characterized by the standard semantics of the logic of here-and-there [Lifschitz et al., 2001]. Such results have rightly been recognized as important for the study of these concrete knowledge representation formalisms, as having a characterization of strong equivalence gives us a deeper level of understanding of the meaning of pieces of knowledge in that formalism.

However, such results about the existence of characterizing formalisms also raise a fundamental question: *Does every formalism have one?* In the following we will see that the answer to this question is a qualified “yes”. More precisely, while not every formalism has one, we show that the important case

of considering only *finite* knowledge bases (but still possibly infinite languages) guarantees the existence of a characterizing formalism, and that in a very general setting. Existing results on characterizing formalisms make use of specifics of each formalism [Lifschitz et al., 2001; Turner, 2001; Truszczyński, 2007; Truszczyński and Woltran, 2008; Cabalar and Diéguez, 2014]. The presented approach completely abstract away from formalism specifics and addresses the core of the problem, the nature of strong equivalence itself. In fact, we will not only show the existence of just any characterizing formalism, but of characterizing formalisms whose model theory is *uniquely determined* (up to isomorphism), and structurally resembles that of classical logics. At this point, we appeal to the reader’s intuition on what makes logics classical; we will later define what we mean by “classical logic” in a precise mathematical way. From our point of view, the main is a surprising and important insight, as it tells us that for the overwhelming majority of knowledge representation formalisms, strong equivalence can be approached using established techniques from classical logic.

While our work is in its essence derived from first principles, building mostly upon classical logic and lattice theory, there have been important inspirations. Foremost, [Truszczyński, 2006] presented a general, algebraic account of strong equivalence within approximation fixpoint theory [Denecker et al., 2000]. His setting is indeed quite general, but most of this generality derives from algebraic commonalities in the semantics of logic programs and default logic. It is not immediately clear, for example, if and how it captures Dung’s theory of abstract argumentation [Dung, 1995], which will be discussed in depth in the subsequent chapters. More precisely, while argumentation frameworks (AFs) with all their semantics can be captured by approximation fixpoint theory [Strass, 2013], Truszczyński’s notion of expanding an operator does not coincide with the corresponding notion of expanding AFs and his results are not directly applicable. In other words, while the operator associated to the union of two logic programs corresponds to the union of their respective associated operators, the same does not hold for the (componentwise) union of two AFs and their operators. Thus although AFs are essentially a restricted subclass of normal logic programs with respect to the ordinary equivalence of having the same models, this does not carry over to strong equivalence because the respective notions of knowledge base union are different in AFs and normal logic programs. For example, the AF where b attacks a corresponds to the logic program $P_1 = \{a \leftarrow \sim b\}$; likewise, $P_2 = \{a \leftarrow \sim c\}$ corresponds to the AF where c attacks a [Osorio et al., 2005; Wu et al., 2009; Strass, 2013]. However, the AF where both b and c attack a (the union of the two AFs above) corresponds to the logic program $P_3 = \{a \leftarrow \sim b, \sim c\}$, where obviously the three programs are not subset-related: $P_1 \not\subseteq P_3$ and $P_2 \not\subseteq P_3$.

In contrast, we show how the approach we develop in this paper can be directly applied to argumentation frameworks. Thus as a consequence of our main theorem, we get the first semantical characterization of strong equivalence in AFs in contrast to the currently known syntactical characterizations [Oikarinen and Woltran, 2011].

The chapter proceeds as follows. In the next section, we introduce the general setting in which we derive our results and present our conception of the term “classical logic”. Afterwards, we define characterization logics and show two classes of formalisms that always possess them. We next apply our results, chiefly to abstract argumentation frameworks but briefly also to normal logic programs. Finally, we conclude the chapter with a discussion of related and future work.

2.1 An Abstract View on Model Theory

What is a classical logic?

We will spend this section introducing an abstract notion of logics with model-theoretic semantics and explaining when we call some of them classical. Formally, we consider logical languages \mathcal{L} , that is, non-empty sets of language elements. We make no assumption on the internal structure of pieces of knowledge $f \in \mathcal{L}$. These pieces of knowledge could be formulas of classical propositional logic, normal

logic program rules, attacks between arguments, or Reiter-style defaults. A model-theoretic semantics for a language \mathcal{L} uses a set I of interpretations and a *model* function $\sigma : 2^{\mathcal{L}} \rightarrow 2^I$ with the intuition that σ assigns each language subset $T \subseteq \mathcal{L}$, a *theory*, the set $\sigma(T)$ of its models. We make no assumptions on the internal structure of interpretations – there need not be an underlying vocabulary of atoms or the like (although in the concrete cases we consider there often will be) that are the same among syntax and semantics. This is the main abstraction in our setting. It goes beyond what is known from classical logic in that meaning is not assigned to language elements (formulas), but only to *theories*, that is, sets of language elements. This is a necessary requirement for being able to model a number of established knowledge representation formalisms: for example, in normal logic programs, meaning is not assigned to single rules, but only to sets thereof. Likewise, in default logic, meaning is not assigned to single defaults, but only to sets thereof. We illustrate our definitions so far by showing more precisely how existing formalisms can be embedded into our setting.

Example 2.1. Consider a set \mathcal{A} of propositional atoms.

Classical propositional logic: The underlying language \mathcal{L}_{PL} is the set of all classical propositional formulas over \mathcal{A} and can be defined as usual by induction. The set of interpretations is then given by the set $I_{PL} = \{v : \mathcal{A} \rightarrow \{\mathbf{t}, \mathbf{f}\}\}$ of all two-valued interpretations of \mathcal{A} . Lastly, $\sigma_{mod}(T)$ is the set of all models of the theory $T \subseteq \mathcal{L}_{PL}$, that is, the set of all interpretations satisfying all formulas in T .

Normal logic programs: The underlying language \mathcal{L}_{LP} is the set of all normal logic program rules $a_0 \leftarrow a_1, \dots, a_m, \sim a_{m+1}, \dots, \sim a_n$ with $0 \leq m \leq n$ and $a_0, a_1, \dots, a_n \in \mathcal{A}$. The set I_{LP} of interpretations is then the set $I_{LP} = 2^{\mathcal{A}}$ of all possible stable model candidates. Accordingly, $\sigma_{stb}(T)$ returns the set of stable models of the theory (normal logic program) $T \subseteq \mathcal{L}_{LP}$ [Gelfond and Lifschitz, 1988]; we could also define $\sigma_{sup}(T)$ returning the set of supported models of T [Clark, 1978], or $\sigma_{mod}(T)$ the set of classical models of T , interpreting \leftarrow as material implication and \sim as classical negation [van Emden and Kowalski, 1976].

Propositional default logic [Reiter, 1980]: The underlying language \mathcal{L}_{DL} is the set of all defaults over formulas over \mathcal{A} , that is, $\mathcal{L}_{DL} = \{(\varphi, \Psi, \xi) \mid \varphi, \xi \in \mathcal{L}_{PL}, \Psi \subseteq \mathcal{L}_{PL}\}$ where the triple $(\varphi, \{\psi_1, \dots, \psi_n\}, \xi)$ represents the default $\varphi : \psi_1, \dots, \psi_n / \xi$. The possible interpretations are given by $I_{DL} = 2^{\mathcal{L}}$, as each default theory is assigned a set of logical theories called extensions: $\sigma_{DL}(T) = \{E \subseteq \mathcal{L} \mid E \text{ is an extension of } T\}$.

Abstract argumentation frameworks [Dung, 1995]: The language \mathcal{L}_{AF} contains the fundamental building blocks of AFs, i.e. arguments and attacks: $\mathcal{L}_{AF} = \{(\{a\}, \emptyset), (\{a, b\}, \{(a, b)\}) \mid a, b \in \mathcal{A}\}$. Extension-based semantics can be incorporated by setting $I_{AF} = 2^{\mathcal{A}}$ and, depending on the argumentation semantics ρ we use, we set $\sigma_\rho(T) = \rho(F_T)$, where $F_T = (\bigcup_{(A,R) \in T} A, \bigcup_{(A,R) \in T} R)$ represents the AF associated to $T \subseteq \mathcal{L}_{AF}$ (cf. Section 2.3.1 for more detailed information and discussion).

Before we delve into the main aim, namely characterizing strong equivalence, we briefly analyze some foundational properties of our way of abstractly modeling knowledge representation languages. We begin with the relationship between the model function and the consequence function of a logical language.

2.1.1 Models and Consequences

A consequence function for a language \mathcal{L} is a function $Cn : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ that assigns a given set T of language elements another set $Cn(T)$ of language elements. Intuitively, $Cn(T)$ is understood to be the set of logical consequences of the theory T . Given a language, we can define the consequence function in

terms of the semantics. In words, the set of consequences of a given theory T is the union of all theories S such that any model of T is a model of S .

Definition 2.2. Let \mathcal{L} be a language and $\sigma : 2^{\mathcal{L}} \rightarrow 2^I$ be a model function. The *canonical consequence function of σ* is defined as follows:

$$Cn^\sigma : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}, \quad T \mapsto \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma(T) \subseteq \sigma(S)}} S$$

In classical definitions of logical consequence, one typically says that a single formula is a consequence of a theory if all models of the theory are models of the formula. In our case, the focus is primarily on theories both for presumptions and consequences, so we stick to the above definition. For classical logic \mathcal{L}_{PL} , this definition coincides with the standard notion of logical consequence.

It will be of great interest that certain algebraic properties of the semantics induce certain useful properties of the consequence relation. We now introduce the most important properties.

Definition 2.3. Let \mathcal{L} be a language.

- A model function $\sigma : 2^{\mathcal{L}} \rightarrow 2^I$ is *antimonotone* iff
for all $T_1, T_2 \in 2^{\mathcal{L}} : T_1 \subseteq T_2 \implies \sigma(T_2) \subseteq \sigma(T_1)$.
- A consequence operator $Cn : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ is *monotone* iff
for all $T_1, T_2 \in 2^{\mathcal{L}} : T_1 \subseteq T_2 \implies Cn(T_1) \subseteq Cn(T_2)$.
- A consequence operator $Cn : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ is *increasing* iff
for all $T \in 2^{\mathcal{L}}$, we find $T \subseteq Cn(T)$.
- A consequence operator $Cn : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ is *idempotent* iff
for all $T \in 2^{\mathcal{L}}$, we find $Cn(Cn(T)) \subseteq Cn(T)$.
- A consequence operator $Cn : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ is a *closure operator* iff
 Cn is monotone, increasing, and idempotent.

In what follows, we define a *logic* as a tuple (\mathcal{L}, I, σ) consisting of a language \mathcal{L} , an interpretation set I , and a model function $\sigma : 2^{\mathcal{L}} \rightarrow 2^I$. We will sometimes associate the canonical consequence function of σ to the whole logic for convenience.

To start out in analyzing how semantics and consequence relate in our setting, we show some straightforward properties of the canonical consequence function defined above. More precisely, any induced consequence operator is increasing and moreover, it is antimonotone if the considered model function is antimonotone.

Proposition 2.4. Let (\mathcal{L}, I, σ) be a logic and Cn^σ its canonical consequence function.

1. Cn^σ is increasing.
2. If σ is antimonotone, then Cn^σ is monotone.

Proof.

1. Clearly for any $T \subseteq \mathcal{L}$ we find $\sigma(T) \subseteq \sigma(T)$ whence $T \subseteq \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma(T) \subseteq \sigma(S)}} S = Cn^\sigma(T)$.
2. Let σ be antimonotone and $T_1 \subseteq T_2$. Then clearly $\sigma(T_2) \subseteq \sigma(T_1)$ by antimonotonicity. Thus for any $S \subseteq \mathcal{L}$ with $\sigma(T_1) \subseteq \sigma(S)$ we also find $\sigma(T_2) \subseteq \sigma(S)$. It follows that

$$\bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma(T_1) \subseteq \sigma(S)}} S \subseteq \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma(T_2) \subseteq \sigma(S)}} S,$$

that is, $Cn^\sigma(T_1) \subseteq Cn^\sigma(T_2)$. Since T_1 and T_2 were arbitrary, Cn^σ is monotone.

□

The reverse of Item 2 does not hold, that is, monotone consequence functions do not necessitate antimonotone model functions.

Example 2.5. Consider $\mathcal{L} = \{a\}$ and $I = \{1\}$ with $\sigma(\emptyset) = \emptyset$ and $\sigma(\{a\}) = \{1\}$. We get the following canonical consequence function:

$$\begin{aligned} Cn^\sigma(\emptyset) &= \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma(\emptyset) \subseteq \sigma(S)}} S = \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \emptyset \subseteq \sigma(S)}} S = \bigcup_{S \subseteq \mathcal{L}} S = \mathcal{L} = \{a\} \\ Cn^\sigma(\{a\}) &= \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma(\{a\}) \subseteq \sigma(S)}} S = \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \{1\} \subseteq \sigma(S)}} S = \{a\} \end{aligned}$$

Thus $Cn^\sigma(\emptyset) = \{a\} = Cn^\sigma(\{a\})$ and Cn^σ is monotone and increasing albeit σ is not antimonotone.

Also, not every antimonotone model function induces a closure operator, that is, an operator that is monotone, increasing and idempotent.

Example 2.6. Consider the logic (\mathcal{L}, I, σ) with language $\mathcal{L} = \{a, b, c\}$, interpretation set $I = \{1\}$ and model function $\sigma : 2^\mathcal{L} \rightarrow 2^I$ given by

$$\sigma(T) = \begin{cases} \{1\} & \text{if } T \in \{\emptyset, \{a\}, \{b\}\} \\ \emptyset & \text{otherwise} \end{cases}$$

In addition to σ being antimonotone we have $Cn^\sigma(\emptyset) = \{a, b\}$ and $Cn^\sigma(\{a, b\}) = \{a, b, c\}$, whence

$$\{a, b, c\} = Cn^\sigma(\{a, b\}) = Cn^\sigma(Cn^\sigma(\emptyset)) \not\subseteq Cn^\sigma(\emptyset) = \{a, b\}$$

which shows that Cn^σ is not idempotent.

We can show that our restriction to semantics via model functions is not overly limiting. We could have chosen to start out from consequence functions as well, as long as these consequence functions are increasing (what is contained in a theory follows from it) and idempotent (all consequences are obtained in one step). More precisely, when given a consequence operator satisfying these two requirements, we can also define a canonical model function whose associated canonical consequence function is exactly the given consequence function we started with.

Proposition 2.7. *Let \mathcal{L} be a language and $C : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ be a consequence function which is increasing and idempotent.*

Then the model function $\sigma_C : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ with $T \mapsto \mathcal{L} \setminus C(T)$ is such that $Cn^{\sigma_C} = C$.¹

Proof. Let $T \subseteq \mathcal{L}$. We find that

$$\begin{aligned}
 Cn^{\sigma_C}(T) &= \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma_C(T) \subseteq \sigma_C(S)}} S && (\text{Def. } Cn^{\sigma}) \\
 &= \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \mathcal{L} \setminus C(T) \subseteq \mathcal{L} \setminus C(S)}} S && (\text{Def. } \sigma_C) \\
 &= \bigcup_{\substack{S \subseteq \mathcal{L}, \\ C(S) \subseteq C(T)}} S && (\text{set algebra}) \\
 &= \bigcup_{\substack{S \subseteq \mathcal{L}, \\ S \subseteq C(S) \subseteq C(T)}} S && (C \text{ is increasing})
 \end{aligned}$$

Firstly, this shows that $Cn^{\sigma_C}(T) \subseteq C(T)$. Moreover, in the last equation we can substitute $S = C(T)$ to obtain that $C(T) \subseteq Cn^{\sigma_C}(T)$. \square

It is clear from Item 1 of Proposition 2.4 that no non-increasing consequence function C can be mimicked by Cn^{σ} for any σ . However, we consider the restrictions of possible consequence functions C having to be increasing and idempotent not to be too severe.

2.1.2 Standard and strong equivalence

This paper is chiefly about characterizing strong equivalence in one logic via standard equivalence in another logic. We will now formally introduce these concepts.

Definition 2.8. Let (\mathcal{L}, I, σ) be a logic and $T_1, T_2 \subseteq \mathcal{L}$ theories. We say that T_1 and T_2 are

- *ordinarily equivalent* if and only if $\sigma(T_1) = \sigma(T_2)$;
- *strongly equivalent* if and only if $\forall U \subseteq \mathcal{L} : \sigma(T_1 \cup U) = \sigma(T_2 \cup U)$.

The notion of strong equivalence is intimately connected with the possibility to simplify parts of a given theory without affecting its semantics. Consider the following example.

Example 2.9. Given a logic (\mathcal{L}, I, σ) , a theory S and a subtheory T_1 of it, i.e. $T_1 \subseteq S$. Now, we may replace T_1 with any T_2 being strongly equivalent to it without changing the semantics of S . More precisely, $\sigma(S) = \sigma(S[T_1|T_2])$ with $S[T_1|T_2] = T_2 \cup (S \setminus T_1)$. This can be seen as follows: Since $T_1 \subseteq S$ we have $T_1 \cup (S \setminus T_1) = S$. Moreover, due to the assumed strong equivalence of T_1 and T_2 we obtain $\sigma(T_1 \cup (S \setminus T_1)) = \sigma(T_2 \cup (S \setminus T_1))$. Hence, $\sigma(S) = \sigma(S[T_1|T_2])$ as claimed.

Clearly, strongly equivalent theories are ordinarily equivalent by definition. What about the converse direction? It is a matter of fact that in case of well-known nonmonotonic formalisms, such as logic programs [Lifschitz et al., 2001], default logic [Turner, 2001], causal theories [Turner, 2004] and abstract argumentation [Oikarinen and Woltran, 2011; Baumann, 2016] strong equivalence and ordinary equivalence are indeed different concepts. However, there are logics like propositional logic or first order logic

¹Please note that the assigned interpretation set I of σ_C is just the language \mathcal{L} .

where both concepts coincide. In the following we will say that the model function σ has the *replacement property* if ordinary equivalence implies strong equivalence. The following natural question arises: What properties must a logic possess in order for ordinary and strong equivalence to coincide? Propositional as well as first order logic possess a monotone consequence function. Does monotony of the consequence operator ensure the coincidence of both concepts? The following example provides us with a negative answer.

Example 2.10. Consider the language $\mathcal{L} = \{a, b\}$ with interpretation set $I = \{1, 2\}$ and model function σ given by

$$\begin{aligned}\sigma(\emptyset) &= \{1, 2\} \\ \sigma(\{a\}) &= \{1, 2\} \\ \sigma(\{b\}) &= \{2\} \\ \sigma(\{a, b\}) &= \emptyset\end{aligned}$$

It is easy to verify that the semantics σ is antimonotone. Therefore, by Proposition 2.4, its consequence function Cn^σ is monotone. However, while \emptyset and $\{a\}$ are obviously ordinarily equivalent, they are not strongly equivalent, which can be seen by extending both with the theory $\{b\}$:

$$\sigma(\emptyset \cup \{b\}) = \sigma(\{b\}) = \{2\} \neq \emptyset = \sigma(\{a, b\}) = \sigma(\{a\} \cup \{b\})$$

We also inspect the induced consequence operator:

$$\begin{aligned}Cn^\sigma(\emptyset) &= \bigcup \{S \subseteq \mathcal{L} \mid \{1, 2\} \subseteq \sigma(S)\} = \{a\} \\ Cn^\sigma(\{a\}) &= \bigcup \{S \subseteq \mathcal{L} \mid \{1, 2\} \subseteq \sigma(S)\} = \{a\} \\ Cn^\sigma(\{b\}) &= \bigcup \{S \subseteq \mathcal{L} \mid \{2\} \subseteq \sigma(S)\} = \{a, b\} \\ Cn^\sigma(\{a, b\}) &= \bigcup \{S \subseteq \mathcal{L} \mid \emptyset \subseteq \sigma(S)\} = \{a, b\}\end{aligned}$$

Since the codomain of Cn^σ consists entirely of fixpoints, Cn^σ is idempotent. Therefore the induced consequence operator Cn^σ is increasing, monotone, and idempotent, thus a closure operator. Yet, the inducing semantics does not have the replacement property.

So having a monotone consequence function is, by itself, insufficient to guarantee the replacement property. We can however identify a property that is strong enough to guarantee replacement on its own. We call it the intersection property, because it basically says that the semantics of a theory can be obtained by only considering the semantics of the singleton sets constituting the theory.

Definition 2.11. Let (\mathcal{L}, I, σ) be a logic. Its model function $\sigma : 2^\mathcal{L} \rightarrow 2^I$ has the *intersection property* iff for all $T \subseteq \mathcal{L}$:

$$\sigma(T) = \bigcap_{F \in T} \sigma(\{F\})$$

It follows from the definition that in particular for any two theories $T_1, T_2 \subseteq \mathcal{L}$, we have that $\sigma(T_1 \cup T_2) = \sigma(T_1) \cap \sigma(T_2)$. The intersection property is a certain locality, independence, or compositionality criterion. Towards an explanation of Example 2.10 we can now remark that its model function σ does not have the intersection property:

$$\sigma(\{a, b\}) = \emptyset \neq \{2\} = \{1, 2\} \cap \{2\} = \sigma(\{a\}) \cap \sigma(\{b\})$$

Indeed, this is necessarily so: as we will show next (and as is easy to show), satisfying the intersection property is sufficient for satisfying the replacement property.

Proposition 2.12. *Let (\mathcal{L}, I, σ) be a logic. If σ satisfies the intersection property, then standard equivalence coincides with strong equivalence.*

Proof. Let σ satisfy the intersection property. It is clear that strong equivalence implies ordinary equivalence (set $U = \emptyset$), so it remains to show the converse. Let $T_1, T_2 \subseteq \mathcal{L}$ such that $\sigma(T_1) = \sigma(T_2)$ and consider any $U \subseteq \mathcal{L}$. We have

$$\begin{aligned} \sigma(T_1 \cup U) &= \sigma(T_1) \cap \sigma(U) && \text{(intersection)} \\ &= \sigma(T_2) \cap \sigma(U) && \text{(presumption)} \\ &= \sigma(T_2 \cup U) && \text{(intersection)} \end{aligned}$$

□

Notably, monotonicity properties were not even needed in the above result. So why is it that all formalisms we know of that have the replacement property also happen to have monotone consequence functions? It holds because σ having the intersection property implies that σ is antimonotone (which in turn implies that Cn^σ is monotone).

Proposition 2.13. *Let (\mathcal{L}, I, σ) be a logic. If σ has the intersection property, then σ is antimonotone.*

Proof. Let $T_1 \subseteq T_2 \subseteq \mathcal{L}$. Then $T_1 \cup T_2 = T_2$, and we conclude the desired subset-inclusion via $\sigma(T_2) = \sigma(T_1 \cup T_2) = \sigma(T_1) \cap \sigma(T_2) \subseteq \sigma(T_1)$. □

In the other direction, we can observe that the replacement property on its own does not guarantee antimonotonicity.

Example 2.14. Consider the language $\mathcal{L} = \{a\}$ and interpretation set $I = \{1\}$. For semantics σ with $\sigma(\emptyset) = \emptyset$ and $\sigma(\{a\}) = \{1\}$, we can see that the replacement property holds trivially since there are no semantically equivalent theories that are syntactically different. However, σ is not antimonotone (as $\sigma(\{a\}) = \{1\} \not\subseteq \emptyset = \sigma(\emptyset)$) and does not have the intersection property:

$$\sigma(\emptyset \cup \{a\}) = \sigma(\{a\}) = \{1\} \neq \emptyset = \emptyset \cap \{1\} = \sigma(\emptyset) \cap \sigma(\{a\})$$

It is easy to see that classical propositional logic \mathcal{L}_{PL} has the intersection property simply by *definition*: the standard model semantics is typically firstly defined for single formulas $\varphi \in \mathcal{L}_{PL}$ and then generalized to theories T by setting $\sigma_{mod}(T) = \bigcap_{\varphi \in T} \sigma_{mod}(\{\varphi\})$.

As it turns out, for all logics, the intersection property also guarantees that each theory T has the same models as the set of all canonical consequences of T .

Proposition 2.15. *Let (\mathcal{L}, I, σ) be a logic that has the intersection property. Then for each $T \subseteq \mathcal{L}$ we find that $\sigma(T) = \sigma(Cn^\sigma(T))$.*

Proof.

$$\begin{aligned} \sigma(Cn^\sigma(T)) &= \sigma\left(\bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma(T) \subseteq \sigma(S)}} S\right) && \text{(Definition 2.2)} \\ &= \bigcap_{\substack{S \subseteq \mathcal{L}, \\ \sigma(T) \subseteq \sigma(S)}} \sigma(S) && \text{(intersection property)} \\ &= \sigma(T) \end{aligned}$$

In the last equality, the \supseteq -direction is clear as we intersect only supersets of $\sigma(T)$, and the \subseteq -direction is clear as we can substitute T for S in the line above. □

Finally, this means that the intersection property only holds for semantics whose canonical consequence functions are closure operators.

Proposition 2.16. *Let (\mathcal{L}, I, σ) be a logic that has the intersection property. Then Cn^σ is a closure operator.*

Proof. We have to show that Cn^σ is increasing, monotone and idempotent. First, Cn^σ is increasing in any case (Proposition 2.4). Moreover, since σ has the intersection property, it is also antimonotone (Proposition 2.13) and thus, Cn^σ is monotone (Proposition 2.4). Finally, it follows from Proposition 2.15 that Cn^σ is also idempotent. More precisely, if $\varphi \in Cn^\sigma(Cn^\sigma(T))$ then there is an $S \subseteq \mathcal{L}$ with $\varphi \in S$ and $\sigma(Cn^\sigma(T)) \subseteq \sigma(S)$, and by $\sigma(T) = \sigma(Cn^\sigma(T))$ it is clear that in this case $\sigma(T) \subseteq \sigma(S)$ and $\varphi \in S \subseteq Cn^\sigma(T)$. Hence, Cn^σ is a closure operator. \square

We have shown in Proposition 2.7 that we also could have started with a consequence function. Clearly the replacement property could be easily defined for consequence functions $C : 2^\mathcal{L} \rightarrow 2^\mathcal{L}$ in the sense that C has the replacement property if and only if for all theories $T_1, T_2 \subseteq \mathcal{L}$, “classical consequence-equivalence” $C(T_1) = C(T_2)$ coincides with “strong consequence-equivalence” $\forall U \subseteq \mathcal{L} : C(T_1 \cup U) = C(T_2 \cup U)$. It is also clear that any semantics having the replacement property induces a canonical consequence function having the consequence-function version of the replacement property. However, we must remark that we have not found a consequence-function equivalent of the intersection property. Even setting $C(T) = \bigcup_{F \in T} C(\{F\})$ would be too weak to capture interactions between different subtheories of T . On the other hand, and on the positive side, we will see that the intersection property for model functions as we defined it in Definition 2.11 will give us a good handle on characterizing strong equivalence.

2.1.3 Galois correspondences

Up to here, we have considered various properties of logics in our abstract setting. Mostly, those were algebraic properties of the model-theoretic semantics. In this subsection, we conclude our argument for defining as “classical” those logics whose semantics satisfy the intersection property.

For this, it is firstly necessary to slightly extend (and, for the time being, also slightly constrain) our abstract notion of “logic”. Up to now, we only assumed the existence of a model function $\sigma : 2^\mathcal{L} \rightarrow 2^I$ that assigns a set of interpretations to a theory (intuitively, its models). In the converse direction, we now also assume a *theory function* $\tau : 2^I \rightarrow 2^\mathcal{L}$, that takes as input a set $K \subseteq I$ of interpretations and intuitively returns the set $\tau(K)$ of all language elements that are true under all interpretations in K . A *logic* will now be a tuple $(\mathcal{L}, I, \sigma, \tau)$ with \mathcal{L}, I, σ as before and $\tau : 2^I \rightarrow 2^\mathcal{L}$ a theory function.

This immediately yields another way of defining a consequence operator: given σ and τ , we can define $Cn^{\sigma, \tau}(T) = \tau(\sigma(T))$. Symmetrically, we can define an operator on interpretation sets by $K \mapsto \sigma(\tau(K))$.

We next consider a class of logics where the interplay of model function and theory function satisfies certain conditions. Below, we denote the composition of functions $f : A \rightarrow B$ and $g : B \rightarrow C$ by $g \circ f$, that is, $g \circ f : A \rightarrow C$ with $x \mapsto g(f(x))$.

Definition 2.17. Let $(\mathcal{L}, I, \sigma, \tau)$ be a logic. The functions σ and τ are in *Galois correspondence* if and only if:

1. σ is antimonotone and τ is antimonotone;
2. $\sigma \circ \tau$ is increasing and $\tau \circ \sigma$ is increasing.

If σ and τ are in Galois correspondence, we also say that $(\mathcal{L}, I, \sigma, \tau)$ is a *Galois logic*.

The first two properties together imply that the resulting consequence function is monotone; by the last property, it is also increasing. Galois correspondences have been studied in model theory [Goguen and Burstall, 1984], and in an even more abstract way in lattice theory [Ore, 1944; Birkhoff, 1973; Cohn, 1981; Davey and Priestley, 2002].

Indeed, there is also another (equivalent) formulation [Davey and Priestley, 2002].

Proposition 2.18. *Let $(\mathcal{L}, I, \sigma, \tau)$ be a logic. It holds that σ and τ are in Galois correspondence iff for all $T \subseteq \mathcal{L}$ and $K \subseteq I$:*

$$K \subseteq \sigma(T) \iff T \subseteq \tau(K)$$

Proof. if: Assume that for all $T \subseteq \mathcal{L}$ and $K \subseteq I$ we have, $K \subseteq \sigma(T)$ iff $T \subseteq \tau(K)$.

1. We start with showing that $\tau \circ \sigma$ is increasing.
Let $T \subseteq \mathcal{L}$. Obviously, $\sigma(T) \subseteq \sigma(T)$. Hence, if substituting $K = \sigma(T)$ we obtain by presumption $T \subseteq \tau(\sigma(T))$.
In the same spirit one may easily show that $\sigma \circ \tau$ is increasing.
2. We show now that σ is antimonotone.
Let $T_1 \subseteq T_2$. Then $T_1 \subseteq T_2 \subseteq \tau(\sigma(T_2))$ by the above. By presumption (with $K = \sigma(T_2)$), it follows that $\sigma(T_2) \subseteq \sigma(T_1)$.
Analogously one may show that τ is antimonotone.

only if: Let $T \subseteq \mathcal{L}$ and $K \subseteq I$. If $K \subseteq \sigma(T)$, then $\tau(\sigma(T)) \subseteq \tau(K)$, whence we conclude that $T \subseteq \tau(\sigma(T)) \subseteq \tau(K)$. The reverse implication follows symmetrically. \square

It also follows that Galois correspondences induce closure operators, that is, operators that are monotone, increasing, and idempotent.

Proposition 2.19. *Let $(\mathcal{L}, I, \sigma, \tau)$ be a Galois logic. Then the operators*

- $\sigma \circ \tau : 2^I \rightarrow 2^I, K \mapsto \sigma(\tau(K))$
- $\tau \circ \sigma : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}, T \mapsto \tau(\sigma(T))$

are closure operators.

Proof. We have to show that $\sigma \circ \tau$ and $\tau \circ \sigma$ are (1) increasing, (2) monotone and (3) idempotent. We only consider $\sigma \circ \tau$ as the proof for $\tau \circ \sigma$ is absolutely symmetric.

1. The second item of Definition 2.17 states verbatim that $\sigma \circ \tau$ is increasing.
2. It is easy to show that $\sigma \circ \tau$ is monotone since both σ and τ are antimonotone: if $K_1 \subseteq K_2$, then $\tau(K_2) \subseteq \tau(K_1)$ whence $\sigma(\tau(K_1)) \subseteq \sigma(\tau(K_2))$, which means $(\sigma \circ \tau)(K_1) \subseteq (\sigma \circ \tau)(K_2)$.
3. Consider $K \subseteq I$. We have to show that $(\sigma \circ \tau)((\sigma \circ \tau)(K)) \subseteq (\sigma \circ \tau)(K)$, that is, $\sigma(\tau(\sigma(\tau(K)))) \subseteq \sigma(\tau(K))$.
Since $\tau \circ \sigma$ is increasing by definition, we have that $\tau(K) \subseteq \tau(\sigma(\tau(K)))$. Since σ is antimonotone, $\sigma(\tau(\sigma(\tau(K)))) \subseteq \sigma(\tau(K))$.

\square

Towards showing that having the intersection property and being in a Galois correspondence are one and the same, we firstly derive a slightly modified characterization of the intersection property. Instead of decomposing theories into its singletons, this version considers families of theories.

Proposition 2.20. *Let (\mathcal{L}, I, σ) be a logic. Its model function $\sigma : 2^{\mathcal{L}} \rightarrow 2^I$ satisfies the intersection property if and only if for all $\mathcal{T} \subseteq 2^{\mathcal{L}}$:*

$$\sigma\left(\bigcup_{T \in \mathcal{T}} T\right) = \bigcap_{T \in \mathcal{T}} \sigma(T)$$

Proof. if: Let $T \subseteq \mathcal{L}$ and define $\mathcal{T} = \{\{F\} \mid F \in T\}$. Clearly $T = \bigcup_{F \in T} \{F\}$, whence

$$\begin{aligned} \sigma(T) &= \sigma\left(\bigcup_{F \in T} \{F\}\right) \\ &= \sigma\left(\bigcup_{\{F\} \in \mathcal{T}} \{F\}\right) \\ &= \bigcap_{\{F\} \in \mathcal{T}} \sigma(\{F\}) \\ &= \bigcap_{F \in T} \sigma(\{F\}) \end{aligned}$$

only if: Let $\mathcal{T} \subseteq 2^{\mathcal{L}}$ and define $T = \bigcup_{U \in \mathcal{T}} U$. Now we have that

$$\begin{aligned} \sigma\left(\bigcup_{U \in \mathcal{T}} U\right) &= \sigma(T) \\ &= \bigcap_{F \in T} \sigma(\{F\}) \\ &= \bigcap_{F \in \bigcup_{U \in \mathcal{T}} U} \sigma(\{F\}) \\ &= \bigcap_{U \in \mathcal{T}} \left(\bigcap_{F \in U} \sigma(\{F\}) \right) \\ &= \bigcap_{U \in \mathcal{T}} \sigma(U) \end{aligned}$$

□

In what follows, we will make implicit use of this result and consider the two formulations of the intersection property to be interchangeable.

We now conclude this section with its main result. It states that for any logic (\mathcal{L}, I, σ) , the conditions “ σ has the intersection property” and “ σ is in a Galois correspondence with some τ ” are equivalent.

The proof can be adapted from the literature [Davey and Priestley, 2002, Propositions 7.31 and 7.33] to our slightly different setting with acceptable effort.

Theorem 2.21. *Let (\mathcal{L}, I, σ) be a logic.*

1. *If there is a $\tau : 2^I \rightarrow 2^{\mathcal{L}}$ such that σ and τ are in Galois correspondence, then σ has the intersection property.*

2. If σ has the intersection property, then we can define a theory function $\tau : 2^I \rightarrow 2^{\mathcal{L}}$ with

$$K \mapsto \bigcup_{\substack{T \subseteq \mathcal{L}, \\ K \subseteq \sigma(T)}} T$$

such that σ and τ are in Galois correspondence.

Proof. 1. Assume the presumption. We have to show that for any $\mathcal{T} \subseteq 2^{\mathcal{L}}$, we find that

$$\sigma\left(\bigcup_{T \in \mathcal{T}} T\right) = \bigcap_{T \in \mathcal{T}} \sigma(T)$$

Let $Z = \bigcup_{T \in \mathcal{T}} T$. We first show that $\sigma(Z)$ is a lower bound for the set $\sigma(\mathcal{T}) = \{\sigma(T) \mid T \in \mathcal{T}\}$. Clearly, for each $T \in \mathcal{T}$ we have $T \subseteq Z$, whence by antimonotonicity of σ we get $\sigma(Z) \subseteq \sigma(T)$ for each $T \in \mathcal{T}$. Now let $Q \subseteq I$ be any lower bound for $\sigma(\mathcal{T})$. Then $Q \subseteq \sigma(T)$ for all $T \in \mathcal{T}$. By Proposition 2.18, we get $T \subseteq \tau(Q)$ for all $T \in \mathcal{T}$. By definition, this entails that $\bigcup_{T \in \mathcal{T}} T = Z \subseteq \tau(Q)$. Now using Proposition 2.18 again yields $Q \subseteq \sigma(Z)$ whence $\sigma(Z)$ is the greatest lower bound of $\sigma(\mathcal{T})$.

2. Let σ have the intersection property. By Proposition 2.13, it follows that σ is antimonotone. For antimonotonicity of τ , consider $K_1 \subseteq K_2$. Then

$$\tau(K_2) = \bigcup_{\substack{T \subseteq \mathcal{L}, \\ K_2 \subseteq \sigma(T)}} T \subseteq \bigcup_{\substack{T \subseteq \mathcal{L}, \\ K_1 \subseteq \sigma(T)}} T = \tau(K_1)$$

Now for showing that $\sigma \circ \tau$ and $\tau \circ \sigma$ are increasing.

Let $K \subseteq I$. We find

$$K \subseteq \bigcap_{\substack{T \subseteq \mathcal{L}, \\ K \subseteq \sigma(T)}} \sigma(T) = \sigma\left(\bigcup_{\substack{T \subseteq \mathcal{L}, \\ K \subseteq \sigma(T)}} T\right) = \sigma(\tau(K))$$

Now let $T \subseteq \mathcal{L}$. By using $\sigma(T) \subseteq \sigma(T)$ it is easy to verify that

$$T \subseteq \bigcup_{\substack{S \subseteq \mathcal{L}, \\ \sigma(T) \subseteq \sigma(S)}} S = \tau(\sigma(T))$$

□

This is the main motivation for our definition saying that *classical* logics are exactly those that have the intersection property, or, equivalently, that have model and theory functions that are in Galois correspondence. Furthermore, as partly mentioned before, many well-studied logics (that we would call classical due to their ubiquity alone) *have* the intersection property simply by definition, as their fundamental building blocks are *formulas* instead of theories.

2.2 Characterization Logics

From now on we omit I from the presentation of logics and thus write (\mathcal{L}, σ) , since concrete interpretations are immaterial for strong equivalence. Furthermore, we will distinguish between two important cases regarding the domain of σ . The first one is $\text{dom}(\sigma) = 2^{\mathcal{L}}$ (called *full logics*) and the second one is $\text{dom}(\sigma) = (2^{\mathcal{L}})_{\text{fin}} = \{T \in 2^{\mathcal{L}} \mid T \text{ is finite}\}$ (*finite-theory logics*), the restriction of \mathcal{L} to finite knowledge bases.

Definition 2.22. Let (\mathcal{L}, σ) be a logic. The binary relation *strong equivalence* $\equiv_s^\sigma \subseteq \text{dom}(\sigma) \times \text{dom}(\sigma)$ is defined by $T_1 \equiv_s^\sigma T_2 \iff \forall U \in \text{dom}(\sigma) : \sigma(T_1 \cup U) = \sigma(T_2 \cup U)$.

It is straightforward to show that \equiv_s^σ is an equivalence relation; we denote the equivalence class of a theory $T \in \text{dom}(\sigma) \subseteq 2^{\mathcal{L}}$ by $[T]_s^\sigma$. We recall and will pervasively use that for all theories $T_1, T_2 \subseteq \mathcal{L}$, we have $T_1 \in [T_2]_s^\sigma$ if and only if $[T_1]_s^\sigma = [T_2]_s^\sigma$.

Given an arbitrary logic (\mathcal{L}, σ) , we want to find a characterizing classical logic, that is, a semantics σ' that has the intersection property and whose ordinary equivalence coincides with strong σ -equivalence. Such logics get a name.

Definition 2.23. Let (\mathcal{L}, σ) be a (full) logic. The logic (\mathcal{L}, σ') is a (full) *characterization logic* for (\mathcal{L}, σ) if and only if:

$$\forall T_1, T_2 \subseteq \mathcal{L} : \sigma'(T_1) = \sigma'(T_2) \iff [T_1]_s^\sigma = [T_2]_s^\sigma \quad (\text{characterization})$$

$$\forall \mathcal{T} \subseteq 2^{\mathcal{L}} : \sigma' \left(\bigcup_{T \in \mathcal{T}} T \right) = \bigcap_{T \in \mathcal{T}} \sigma'(T) \quad (\text{intersection})$$

Since characterization logics are the centerpiece of our study we would like to take a look to this central definition from an other angle, namely in terms of consequence functions. Note that the analysis of consequence relations have been very prominent in the formative years of nonmonotonic reasoning. We refer the interested reader to [Gabbay, 1985; Kraus et al., 1990; Makinson, 1994]. Let us consider the canonical consequence functions of σ and σ' according to Definition 2.2. Doing so reveals that the characterizing and characterized consequence relations fit together appropriately as stated in the following assertion. It is part of future work to study further properties in terms of consequence relations.

Proposition 2.24. Given a logic (\mathcal{L}, σ) and a characterization logic (\mathcal{L}, σ') of it. Let Cn^σ and $\text{Cn}^{\sigma'}$ be the canonical consequence functions of the corresponding logics. Given $T \subseteq \mathcal{L}$.

1. $\text{Cn}^{\sigma'}(T) \subseteq \text{Cn}^\sigma(T)$ (sublogic)
2. $\text{Cn}^\sigma(T) = \text{Cn}^{\sigma'}(\text{Cn}^\sigma(T))$ (left absorption)
3. $\text{Cn}^\sigma(T) = \text{Cn}^\sigma(\text{Cn}^{\sigma'}(T))$ (right absorption)

Proof. In the following proofs we will often use that (\mathcal{L}, σ) and (\mathcal{L}, σ') are characterization logics implying that σ and σ' possess the intersection property (Definition 2.23) which in turn means that Cn^σ and $\text{Cn}^{\sigma'}$ are closure operators (Proposition 2.16).

1. In order to show $\text{Cn}^{\sigma'}(T) \subseteq \text{Cn}^\sigma(T)$ we first prove the subsequent property (*). For any $U \subseteq \mathcal{L}$: If $U \subseteq \text{Cn}^{\sigma'}(T)$, then $\text{Cn}^{\sigma'}(T) = \text{Cn}^{\sigma'}(T \cup U)$. Since $T \subseteq T \cup U$ we derive $\text{Cn}^{\sigma'}(T) \subseteq \text{Cn}^{\sigma'}(T \cup U)$ (monotonicity). Moreover, we have $T \subseteq \text{Cn}^{\sigma'}(T)$ (increasing) and $U \subseteq \text{Cn}^{\sigma'}(T)$ (assumption) justifying $T \cup U \subseteq \text{Cn}^{\sigma'}(T)$. Hence, $\text{Cn}^{\sigma'}(T \cup U) \subseteq \text{Cn}^{\sigma'}(\text{Cn}^{\sigma'}(T))$ (monotonicity) and finally, $\text{Cn}^{\sigma'}(T \cup U) \subseteq \text{Cn}^{\sigma'}(T)$ (idempotency) concluding the proof for property (*).

Now, let $U \subseteq Cn^{\sigma'}(T)$. We have to show $U \subseteq Cn^{\sigma}(T)$. Due to Proposition 2.15 and property (*) we obtain $\sigma'(T) = \sigma'(Cn^{\sigma'}(T)) = \sigma'(Cn^{\sigma'}(T \cup U)) = \sigma'(T \cup U)$. Since (\mathcal{L}, σ') is assumed to be a characterization logic we derive $[T]_s^{\sigma} = [T \cup U]_s^{\sigma}$. Consequently, $\sigma(T) = \sigma(T \cup U)$ implying $Cn^{\sigma}(T) = Cn^{\sigma}(T \cup U)$ in consideration of Definition 2.2. Using the previous equality justifies $U \subseteq T \cup U \subseteq Cn^{\sigma}(T)$.

2. On the one hand, we have $Cn^{\sigma}(T) \subseteq Cn^{\sigma'}(Cn^{\sigma}(T))$ since $Cn^{\sigma'}$ is increasing. On the other hand, $Cn^{\sigma'}(Cn^{\sigma}(T)) \subseteq Cn^{\sigma}(Cn^{\sigma}(T))$ due to item 1 (sublogic) and finally, $Cn^{\sigma'}(Cn^{\sigma}(T)) \subseteq Cn^{\sigma}(T)$ since Cn^{σ} is increasing too. Thus, $Cn^{\sigma}(T) = Cn^{\sigma'}(Cn^{\sigma}(T))$ as claimed.
3. Due to Proposition 2.15 we have $\sigma'(T) = \sigma'(Cn^{\sigma'}(T))$. Since (\mathcal{L}, σ') is a characterization logic we further conclude $[T]_s^{\sigma} = [Cn^{\sigma'}(T)]_s^{\sigma}$. Hence, $\sigma(T) = \sigma(Cn^{\sigma'}(T))$ which guarantees $Cn^{\sigma}(T) = Cn^{\sigma}(Cn^{\sigma'}(T))$ by Definition 2.2.

□

We now start our analysis of characterization logics in terms of model functions. We first show that characterization logics are unique up to isomorphism. More precisely, for any model function σ , the algebras corresponding to the model theories of any two characterizing model functions σ' and σ'' are isomorphic. To do that, we first show that the model theory of any characterization logic² is a complete lattice, that is, a partially ordered set where each subset of the carrier set has both a greatest lower bound (glb) and a least upper bound (lub).

Theorem 2.25. *Let (\mathcal{L}, σ) be a full logic with characterization logic (\mathcal{L}, σ') . The pair $(\sigma'(2^{\mathcal{L}}), \subseteq)$ is a complete lattice where glb \wedge and lub \vee are given such that for all $\mathcal{K} \subseteq \sigma'(2^{\mathcal{L}})$,*

$$\bigwedge_{K \in \mathcal{K}} K = \bigcap_{K \in \mathcal{K}} K \text{ and } \bigvee_{K \in \mathcal{K}} K = \bigwedge_{L \in \mathcal{K}^u} L$$

where $\mathcal{K}^u = \{L \in \sigma'(2^{\mathcal{L}}) \mid \forall K \in \mathcal{K} : K \subseteq L\}$ is the set of upper bounds of \mathcal{K} .

Proof. Let $\mathcal{K} \subseteq \sigma'(2^{\mathcal{L}})$; we first show $\bigcap_{K \in \mathcal{K}} K \in \sigma'(2^{\mathcal{L}})$. Clearly for each $K \in \mathcal{K} \subseteq \sigma'(2^{\mathcal{L}})$ there exists a $T \subseteq \mathcal{L}$ with $\sigma'(T) = K$. Thus by the axiom of choice there is a $\mathcal{T} \subseteq 2^{\mathcal{L}}$ that contains a $T \in \mathcal{T}$ with $\sigma'(T) = K$ for each $K \in \mathcal{K}$. Since $\bigcup_{T \in \mathcal{T}} T \in 2^{\mathcal{L}}$ and σ' has the intersection property, $\bigcap_{K \in \mathcal{K}} K = \bigcap_{T \in \mathcal{T}} \sigma'(T) = \sigma'(\bigcup_{T \in \mathcal{T}} T) \in \sigma'(2^{\mathcal{L}})$.

Now consider $\bigvee_{K \in \mathcal{K}} K$. We show that $\bigwedge_{L \in \mathcal{K}^u} L$ is the least element of \mathcal{K}^u , the set of all upper bounds of \mathcal{K} . Clearly, $L \in \mathcal{K}^u$ implies that $\forall K \in \mathcal{K} : K \subseteq L$. Thus, $\forall K \in \mathcal{K} : K \subseteq \bigcap_{L \in \mathcal{K}^u} L$ whence $\bigcap_{L \in \mathcal{K}^u} L \in \mathcal{K}^u$. In particular, if $M \in \mathcal{K}^u$ then $\bigcap_{L \in \mathcal{K}^u} L \subseteq M$ and $\bigcap_{L \in \mathcal{K}^u} L$ is the least upper bound of \mathcal{K} . □

It is vital that the least upper bound is defined in terms of the greatest lower bound, as ordinary set union will not work.

Example 2.26. Consider $\mathcal{L} = \{a, b, c\}$ and semantics σ with

$$\begin{aligned} [\emptyset]_s^{\sigma} &= \{\emptyset\}, \\ [\{a\}]_s^{\sigma} &= \{\{a\}\}, \\ [\{b\}]_s^{\sigma} &= \{\{b\}\}, \\ [\{c\}]_s^{\sigma} &= \{\{c\}\}, \\ [\{a, b\}]_s^{\sigma} &= [\{a, c\}]_s^{\sigma} = [\{b, c\}]_s^{\sigma} = [\{a, b, c\}]_s^{\sigma} = \{\{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\} \end{aligned}$$

²The proof of Theorem 2.25 reveals that (\mathcal{L}, σ') does not necessarily has to be a characterization logic of (\mathcal{L}, σ) . Indeed, the stated properties regarding the model theory hold for any logic possessing the intersection property.

Assume that (\mathcal{L}, σ') is a characterization logic for (\mathcal{L}, σ) . We claim that

$$\sigma'(\{b\}) \cup \sigma'(\{c\}) \notin \sigma'(2^{\mathcal{L}})$$

whence $\sigma'(2^{\mathcal{L}})$ is not closed under set union and thus cannot be the carrier set of any sublattice of $(2^{\mathcal{L}}, \subseteq)$. Assume to the contrary that there is a $T \subseteq \mathcal{L}$ with $\sigma'(T) = \sigma'(\{b\}) \cup \sigma'(\{c\})$. Then

$$\begin{aligned} \sigma'(\{a\} \cup T) &= \sigma'(\{a\}) \cap \sigma'(T) \\ &= \sigma'(\{a\}) \cap (\sigma'(\{b\}) \cup \sigma'(\{c\})) \\ &= (\sigma'(\{a\}) \cap \sigma'(\{b\})) \cup (\sigma'(\{a\}) \cap \sigma'(\{c\})) \\ &= \sigma'(\{a, b\}) \cup \sigma'(\{a, c\}) \\ &= \sigma'(\{a, b\}) = \sigma'(\{a, c\}) = \sigma'(\{b, c\}) = \sigma'(\{a, b, c\}) \end{aligned}$$

Thus $T = \{b\}$ or $T = \{c\}$ or $T \in [\{a, b, c\}]_s^\sigma$. Since σ' has the intersection property, it is in particular antimonotone (Proposition 2.13). Hence for $T \in [\{a, b, c\}]_s^\sigma$, we get $\sigma'(T) \subseteq \sigma'(\{b\})$ and since $[\{b\}]_s^\sigma \neq [\{a, b, c\}]_s^\sigma$ even $\sigma'(T) \subsetneq \sigma'(\{b\})$. Substituting c for b in the argument above, we obtain $\sigma'(T) \subsetneq \sigma'(\{c\})$ in the same fashion. In combination $\sigma'(T) \subsetneq \sigma'(\{b\}) \cup \sigma'(\{c\})$, contradiction. Thus $T = \{b\}$ or $T = \{c\}$. Clearly $[\{b\}]_s^\sigma \neq [\{b, c\}]_s^\sigma$ and $[\{c\}]_s^\sigma \neq [\{b, c\}]_s^\sigma$ imply that

$$\begin{aligned} \sigma'(\{b\}) \cap \sigma'(\{c\}) &= \sigma'(\{b, c\}) \subsetneq \sigma'(\{b\}) \\ \sigma'(\{b\}) \cap \sigma'(\{c\}) &= \sigma'(\{b, c\}) \subsetneq \sigma'(\{c\}) \end{aligned}$$

Thus $\sigma'(\{b\}) \setminus \sigma'(\{c\}) \neq \emptyset$ and $\sigma'(\{c\}) \setminus \sigma'(\{b\}) \neq \emptyset$. Therefore

$$\sigma'(\{b\}) \subsetneq \sigma'(\{b\}) \cup \sigma'(\{c\}) = \sigma'(T)$$

as well as

$$\sigma'(\{c\}) \subsetneq \sigma'(\{b\}) \cup \sigma'(\{c\}) = \sigma'(T).$$

Contradiction. Thus T does not exist and $\sigma'(\{b\}) \cup \sigma'(\{c\}) \notin \sigma'(2^{\mathcal{L}})$.

This example shows in particular that the resulting complete lattice induced by the characterization logic is not necessarily distributive, not even in the case of finite logics.

After these necessary preliminaries, we now present the result on uniqueness of characterization logics.

Theorem 2.27. *Let (\mathcal{L}, σ) be a full logic with characterization logics (\mathcal{L}, σ') and (\mathcal{L}, σ'') . Then the complete lattices $(\sigma'(2^{\mathcal{L}}), \subseteq)$ and $(\sigma''(2^{\mathcal{L}}), \subseteq)$ are isomorphic.*

Proof. We provide a bijection $\phi : \sigma'(2^{\mathcal{L}}) \rightarrow \sigma''(2^{\mathcal{L}})$ with

$$\phi\left(\bigwedge_{K \in \mathcal{K}} K\right) = \bigwedge_{K \in \mathcal{K}} \phi(K) \text{ and } \phi\left(\bigvee_{K \in \mathcal{K}} K\right) = \bigvee_{K \in \mathcal{K}} \phi(K)$$

Given any $K \in \sigma'(2^{\mathcal{L}})$, there clearly exists a $T \subseteq \mathcal{L}$ with $\sigma'(T) = K$; now define $\phi(K)$ such that $\phi(K) = \sigma''(T)$. This means, $\phi(\sigma'(T)) = \phi(\sigma''(T))$.

- ϕ is injective: Let $K_1, K_2 \in \sigma'(2^{\mathcal{L}})$ with $\phi(K_1) = \phi(K_2)$. Clearly there exist $T_1, T_2 \subseteq \mathcal{L}$ such that $\sigma'(T_1) = K_1$ and $\sigma'(T_2) = K_2$. Thus $\phi(K_1) = \phi(K_2)$ implies $\sigma''(T_1) = \phi(K_1) = \phi(K_2) = \sigma''(T_2)$. Since σ'' has the characterization property, we get $[T_1]_s^\sigma = [T_2]_s^\sigma$. Since σ' has the characterization property, we get $K_1 = \sigma'(T_1) = \sigma'(T_2) = K_2$.

- ϕ is surjective: Let $M \in \sigma''(2^{\mathcal{L}})$. Then there exists a $T \subseteq \mathcal{L}$ with $\sigma''(T) = M$. It follows by definition that $\phi(\sigma'(T)) = \sigma''(T) = M$.
- ϕ is structure-preserving: Let $\mathcal{K} \subseteq \sigma'(2^{\mathcal{L}})$ and define $\mathcal{T} \subseteq 2^{\mathcal{L}}$ such that for each $K \in \mathcal{K}$, the family \mathcal{T} contains an element of the preimage of K with respect to σ' . (\mathcal{T} need not be unique, but exists by the axiom of choice.)

$$\begin{aligned} \phi\left(\bigwedge_{K \in \mathcal{K}} K\right) &= \phi\left(\bigcap_{K \in \mathcal{K}} K\right) = \phi\left(\bigcap_{T \in \mathcal{T}} \sigma'(T)\right) = \phi\left(\sigma'\left(\bigcup_{T \in \mathcal{T}} T\right)\right) \\ &= \sigma''\left(\bigcup_{T \in \mathcal{T}} T\right) = \bigcap_{T \in \mathcal{T}} \sigma''(T) = \bigcap_{K \in \mathcal{K}} \phi(K) = \bigwedge_{K \in \mathcal{K}} \phi(K) \end{aligned}$$

Since \vee can be defined in terms of \wedge , it follows that

$$\phi\left(\bigvee_{K \in \mathcal{K}} K\right) = \phi\left(\bigwedge_{K \in \mathcal{K}^u} K\right) = \bigwedge_{K \in \mathcal{K}^u} \phi(K) = \bigvee_{K \in \mathcal{K}} \phi(K)$$

Thus $(\sigma'(2^{\mathcal{L}}), \subseteq)$ and $(\sigma''(2^{\mathcal{L}}), \subseteq)$ are isomorphic. \square

Thus if a classical characterization logic exists, it is (up to isomorphism on its model theory) uniquely determined. However, as we show next, in some cases there simply is no characterization logic.

Example 2.28. Let $\mathcal{L} = \mathbb{N}$ be the natural numbers and $I \neq \emptyset$ arbitrary. We define the semantics $\sigma : 2^{\mathcal{L}} \rightarrow 2^I$ such that

$$\sigma(T) = \begin{cases} \emptyset & \text{if } T \text{ is finite,} \\ I & \text{otherwise.} \end{cases}$$

There are two strong equivalence classes: $[\emptyset]_s^\sigma$, the set of all finite subsets of \mathbb{N} , and $[\mathbb{N}]_s^\sigma$, the set of all infinite subsets of \mathbb{N} . Assume that (\mathcal{L}, σ') is a characterization logic for (\mathcal{L}, σ) . By the model intersection property, we get

$$\sigma'(\mathbb{N}) = \sigma'\left(\bigcup_{n \in \mathbb{N}} \{n\}\right) = \bigcap_{n \in \mathbb{N}} \sigma'(\{n\}) = \sigma'(\emptyset)$$

in contradiction to the characterization property.

The example above motivates the study of special features of logics warranting the existence of characterization logics. We now proceed with some useful properties that will be needed in this endeavour later on. Most importantly, we show that strong equivalence classes have an expansion property: It is not completely obvious, but follows easily from the definition of strong equivalence that two strongly equivalent theories can both be expanded (via set union) with the same theory and are again strongly equivalent; the converse holds as well. Furthermore, the union of two strongly equivalent theories is again strongly equivalent to the two theories. Finally, for any two strongly equivalent theories that are in subset relation, every theory in between them is also strongly equivalent to them.

Lemma 2.29. Let (\mathcal{L}, σ) be a full logic and $T, T_1, T_2, T_3 \subseteq \mathcal{L}$.

1. Strong equivalence is invariant to expansion:

$$[T_1]_s^\sigma = [T_2]_s^\sigma \iff \left(\forall U \subseteq \mathcal{L} : [T_1 \cup U]_s^\sigma = [T_2 \cup U]_s^\sigma \right)$$

2. Each strong equivalence class is a join-semilattice:

$$T_1, T_2 \in [T]_s^\sigma \implies T_1 \cup T_2 \in [T]_s^\sigma$$

3. Each strong equivalence class is convex:

$$T_1 \in [T]_s^\sigma \wedge T_3 \in [T]_s^\sigma \wedge T_1 \subseteq T_2 \subseteq T_3 \implies T_2 \in [T]_s^\sigma$$

Proof. 1.

$$\begin{aligned} [T_1]_s^\sigma &= [T_2]_s^\sigma \\ \iff \forall U \subseteq \mathcal{L} : \sigma(T_1 \cup U) &= \sigma(T_2 \cup U) && (\text{Def. } \equiv_s^\sigma) \\ \iff \forall U' \subseteq \mathcal{L} : \forall U'' \subseteq \mathcal{L} : \sigma(T_1 \cup (U' \cup U'')) &= \sigma(T_2 \cup (U' \cup U'')) && (\text{write } U = U' \cup U'') \\ \iff \forall U' \subseteq \mathcal{L} : \forall U'' \subseteq \mathcal{L} : \sigma((T_1 \cup U') \cup U'') &= \sigma((T_2 \cup U') \cup U'') && (\text{associativity } \cup) \\ \iff \forall U' \subseteq \mathcal{L} : [T_1 \cup U']_s^\sigma &= [T_2 \cup U']_s^\sigma && (\text{Def. } \equiv_s^\sigma) \\ \iff \forall U \subseteq \mathcal{L} : [T_1 \cup U]_s^\sigma &= [T_2 \cup U]_s^\sigma && (\text{rename } U' = U) \end{aligned}$$

2.

$$\begin{aligned} T_1, T_2 &\in [T]_s^\sigma \\ \iff [T_1]_s^\sigma &= [T_2]_s^\sigma = [T]_s^\sigma \\ \implies [T_1 \cup T_2]_s^\sigma &= [T_2 \cup T_2]_s^\sigma = [T_2]_s^\sigma = [T]_s^\sigma \\ \implies T_1 \cup T_2 &\in [T]_s^\sigma \end{aligned}$$

3.

$$\begin{aligned} T_1, T_3 &\in [T]_s^\sigma \wedge T_1 \subseteq T_2 \subseteq T_3 \\ \implies (T_1 \cup T_2 \in [T]_s^\sigma &\iff T_3 \cup T_2 \in [T]_s^\sigma) \wedge T_1 \subseteq T_2 \subseteq T_3 \\ \implies (T_2 \in [T]_s^\sigma &\iff T_3 \in [T]_s^\sigma) \\ \implies T_2 &\in [T]_s^\sigma \end{aligned}$$

□

We also take a closer look on the structure of strong equivalence classes and their relationships with each other. We first define a binary relation \sqsubseteq on strong equivalence classes where two classes are in \sqsubseteq -relation if and only if there are \sqsubseteq -comparable representatives of the classes.

Definition 2.30. Let (\mathcal{L}, σ) be a logic. Define the following relation on strong equivalence classes:

$$[S]_s^\sigma \sqsubseteq [T]_s^\sigma \iff \exists S' \in [S]_s^\sigma : \exists T' \in [T]_s^\sigma : S' \subseteq T'$$

It is easy to show (and later useful) that \sqsubseteq is a partial order.

Lemma 2.31. The relation \sqsubseteq is a partial order, that is, reflexive, antisymmetric and transitive.

Proof. Reflexivity is clear.

antisymmetric: Let $[S]_s^\sigma \subseteq [T]_s^\sigma$ and $[T]_s^\sigma \subseteq [S]_s^\sigma$. Then there are $S_1, S_2 \in [S]_s^\sigma$ as well as $T_1, T_2 \in [T]_s^\sigma$ with $S_1 \subseteq T_1$ and $T_2 \subseteq S_2$. Clearly

$$\begin{aligned}
[S]_s^\sigma &= [S_1 \cup S_2]_s^\sigma && (S_1, S_2 \in [S]_s^\sigma; \text{Lemma 2.29, Item 2}) \\
&= [T_2 \cup S_1 \cup S_2]_s^\sigma && (T_2 \subseteq S_2) \\
&= [T_1 \cup T_2 \cup S_1 \cup S_2]_s^\sigma && ([T_2]_s^\sigma = [T_1 \cup T_2]_s^\sigma) \\
&= [T_1 \cup T_2 \cup S_1]_s^\sigma && ([S_1 \cup S_2]_s^\sigma = [S_1]_s^\sigma) \\
&= [T_1 \cup T_2]_s^\sigma && (S_1 \subseteq T_1) \\
&= [T]_s^\sigma && (T_1, T_2 \in [T]_s^\sigma; \text{Lemma 2.29, Item 2})
\end{aligned}$$

transitive: Let $[T_1]_s^\sigma \subseteq [T_2]_s^\sigma$ and $[T_2]_s^\sigma \subseteq [T_3]_s^\sigma$. Then there exist $T'_1 \in [T_1]_s^\sigma$, $T'_2, T''_2 \in [T_2]_s^\sigma$ and $T''_3 \in [T_3]_s^\sigma$ with $T'_1 \subseteq T'_2$ and $T'_2 \subseteq T''_3$. Since $[T'_2]_s^\sigma = [T''_2]_s^\sigma$, we conclude that $[T'_2 \cup T''_3]_s^\sigma = [T''_2 \cup T''_3]_s^\sigma = [T''_3]_s^\sigma$. Now it is clear that $T'_1 \subseteq T'_2 \subseteq T'_2 \cup T''_3$ with $T'_1 \in [T_1]_s^\sigma$ and $T'_2 \cup T''_3 \in [T_3]_s^\sigma$ imply that $[T_1]_s^\sigma \subseteq [T_3]_s^\sigma$. \square

It follows from Lemma 2.29 in particular that in the case of logics (\mathcal{L}, σ) with \mathcal{L} finite, each strong equivalence class $[T]_s^\sigma$ has a \subseteq -greatest element that equals the union of all elements. However, for logics with infinite \mathcal{L} , this need not be the case: in the logic of Example 2.28, the class $[\emptyset]_s^\sigma$ has no maximal elements, in particular no greatest element; the class $[\mathbb{N}]_s^\sigma$ has no minimal elements, in particular no least element. We will see that having a \subseteq -greatest element in each equivalence class is sufficient for the existence of a characterization logic. We therefore decided to name this class of logics and study it in some greater detail.

2.2.1 Covered Logics

Definition 2.32. Let (\mathcal{L}, σ) be a logic with strong equivalence relation \equiv_s^σ . For an equivalence class $[T]_s^\sigma \in (2^\mathcal{L})_{/\equiv_s^\sigma} = \{[T]_s^\sigma \mid T \subseteq \mathcal{L}\}$, we define its *cover* to be the set

$$\widehat{[T]_s^\sigma} = \bigcup_{S \in [T]_s^\sigma} S$$

and say that a logic (\mathcal{L}, σ) is *covered* if and only if $\forall T \subseteq \mathcal{L} : \widehat{[T]_s^\sigma} \in [T]_s^\sigma$.

Roughly, the existence of greatest elements in equivalence classes guarantees that these classes are closed under arbitrary set union.³ Clearly any finite logic is covered. Furthermore, two familiar representatives of covered logics are classical logic and abstract argumentation theory. In the former case, it is clear that arbitrary unions of families of equivalent theories are again theories that are equivalent to each of its members. In the latter case it is not immediately clear but can be shown with reasonable effort.

Towards the main result of this section, we show that covered logics behave “nicely” in an algebraic sense, which will pave the way for obtaining characterization logics for them. Most importantly, while we know from Theorem 2.25 that the strong equivalence classes of any logic form a complete lattice, for covered logics we can even specify the join and meet operations directly via set operations on covers and subsequent class formation.

³Strong equivalence classes are always closed under set unions with non-empty, finite index sets (Lemma 2.29).

Lemma 2.33. *Let (\mathcal{L}, σ) be a covered logic with strong equivalence relation \equiv_s^σ . The pair $((2^\mathcal{L})_{/\equiv_s^\sigma}, \sqsubseteq)$ is a complete lattice with operations*

$$\bigsqcup_{C \in \mathcal{C}} C = \left[\bigcup_{C \in \mathcal{C}} \widehat{C} \right]_s^\sigma \text{ and } \bigcap_{C \in \mathcal{C}} C = \left[\bigcap_{C \in \mathcal{C}} \widehat{C} \right]_s^\sigma$$

Proof. Let $\mathcal{C} \subseteq \{[T]_s^\sigma \mid T \subseteq \mathcal{L}\}$.

- $D = \bigsqcup_{C \in \mathcal{C}} C$ is the least upper bound of \mathcal{C} :

For any $C \in \mathcal{C}$, we get $\widehat{C} \subseteq \bigcup_{B \in \mathcal{C}} \widehat{B}$ immediately and $\widehat{C} \in C$ since (\mathcal{L}, σ) is covered. The fact that $\bigcup_{B \in \mathcal{C}} \widehat{B} \in \left[\bigcup_{B \in \mathcal{C}} \widehat{B} \right]_s^\sigma = D$ is also immediate, whence $C \sqsubseteq D$ and by arbitrary choice of C we get that D is an upper bound of \mathcal{C} .

Now let E be any upper bound of \mathcal{C} and consider an arbitrary $C \in \mathcal{C}$. Since $C \sqsubseteq E$, there are $T_C \in C$ and $T'_C \in E$ such that $T_C \subseteq T'_C$. Since (\mathcal{L}, σ) is covered, $[T_C]_s^\sigma = C = \left[\widehat{C} \right]_s^\sigma$ whence by strong equivalence $[T_C \cup T'_C]_s^\sigma = \left[\widehat{C} \cup T'_C \right]_s^\sigma$. That is, $\widehat{C} \cup T'_C \in [T_C \cup T'_C]_s^\sigma = [T'_C]_s^\sigma = E$. Since C was arbitrary, we have that $\forall C \in \mathcal{C} : \exists T'_C \in E : \widehat{C} \cup T'_C \in E$. In particular, $\forall C \in \mathcal{C} : \exists T'_C \in E : \widehat{C} \cup T'_C \subseteq \widehat{E}$ whence $\bigcup_{C \in \mathcal{C}} (\widehat{C} \cup T'_C) \subseteq \widehat{E}$. Now fix a specific $B \in \mathcal{C}$; then we have in particular that there exists a $T'_B \in E$ such that $\widehat{B} \cup T'_B \in E$. Furthermore, $\widehat{B} \cup T'_B \subseteq \bigcup_{C \in \mathcal{C}} (\widehat{C} \cup T'_C) \subseteq \widehat{E}$ with $\widehat{E} \in E$ since (\mathcal{L}, σ) is covered. By Lemma 2.29 (Item 3) saying that strong equivalence classes are convex, we get $\bigcup_{C \in \mathcal{C}} (\widehat{C} \cup T'_C) \in E$. Together with $\bigcup_{C \in \mathcal{C}} \widehat{C} \subseteq \bigcup_{C \in \mathcal{C}} (\widehat{C} \cup T'_C)$ and $\bigcup_{C \in \mathcal{C}} \widehat{C} \in D$ we get $D \sqsubseteq E$ and D is the least upper bound of \mathcal{C} .

- $D = \bigcap_{C \in \mathcal{C}} C$ is the greatest lower bound of \mathcal{C} :

As above, $C \in \mathcal{C}$ implies $\bigcap_{B \in \mathcal{C}} \widehat{B} \subseteq \widehat{C}$, whence by $\widehat{C} \in C$ and $\bigcap_{B \in \mathcal{C}} \widehat{B} \in D$ we get $D \sqsubseteq C$. Since $C \in \mathcal{C}$ was arbitrarily chosen, D is a lower bound of \mathcal{C} .

Now let E be any lower bound of \mathcal{C} . Thus by definition, for each $C \in \mathcal{C}$ there exist $T_C \in C$ and $T'_C \in E$ such that $T'_C \subseteq T_C$. Now fix one $C \in \mathcal{C}$ and consider an arbitrary $B \in \mathcal{C}$. Clearly $[T'_C]_s^\sigma = [T'_B]_s^\sigma = E$ whence $[T'_C \cup T_B]_s^\sigma = [T'_B \cup T_B]_s^\sigma = [T_B]_s^\sigma$ by strong equivalence and since $T'_B \subseteq T_B$. Therefore, $T'_C \cup T_B \in B$, that is, $T'_C \subseteq T'_C \cup T_B \subseteq \widehat{B}$ with $\widehat{B} \in B$ since (\mathcal{L}, σ) is covered. Since B was chosen arbitrarily, we get the following: for every $B \in \mathcal{C}$, we have that $T'_C \subseteq \widehat{B}$. Consequently $T'_C \subseteq \bigcap_{B \in \mathcal{C}} \widehat{B}$ with $T'_C \in E$ and $\bigcap_{B \in \mathcal{C}} \widehat{B} \in D$. Thus $E \sqsubseteq D$ and D is the greatest lower bound of \mathcal{C} . □

In particular, the least and greatest elements of that lattice are given by

$$\bigsqcup_{C \in \emptyset} C = \left[\bigcup_{C \in \emptyset} \widehat{C} \right]_s^\sigma = [\emptyset]_s^\sigma \text{ and } \bigcap_{C \in \emptyset} C = \left[\bigcap_{C \in \emptyset} \widehat{C} \right]_s^\sigma = [\mathcal{L}]_s^\sigma.$$

As we show next, it follows that the mapping $[\cdot]_s^\sigma : 2^\mathcal{L} \rightarrow (2^\mathcal{L})_{/\equiv_s^\sigma}$ assigning a theory T its equivalence class $[T]_s^\sigma$ is join-preserving for arbitrary joins. This is akin to the intersection property for the characterizing semantics, only that the semantics is not yet a model theory in the form of a complete lattice of sets but, more generally, a complete lattice (that is not necessarily one of sets).

Lemma 2.34. Let (\mathcal{L}, σ) be a covered logic with strong equivalence relation \equiv_s^σ . For all $\mathcal{T} \subseteq 2^\mathcal{L}$, we have:

$$\left[\bigcup_{T \in \mathcal{T}} T \right]_s^\sigma = \bigsqcup_{T \in \mathcal{T}} [T]_s^\sigma$$

Proof. Let $\mathcal{T} \subseteq 2^\mathcal{L}$ and denote $\mathcal{C} = \{[T]_s^\sigma \mid T \in \mathcal{T}\}$. Clearly $\bigsqcup_{T \in \mathcal{T}} [T]_s^\sigma = \bigsqcup \mathcal{C}$. Consequently, it suffices to show that

$$\left[\bigcup_{T \in \mathcal{T}} T \right]_s^\sigma = \bigsqcup \mathcal{C}.$$

This will be done by showing that $D = [\bigcup_{T \in \mathcal{T}} T]_s^\sigma$ is the least upper bound of \mathcal{C} in $((2^\mathcal{L})_{/\equiv_s^\sigma}, \sqsubseteq)$.

Let $C \in \mathcal{C}$. Then there is a $T \in \mathcal{T}$ with $C = [T]_s^\sigma$. Clearly $T \in C$ and

$$T \subseteq \bigcup_{S \in \mathcal{T}} S \subseteq \left[\bigcup_{S \in \mathcal{T}} S \right]_s^\sigma = \widehat{D}$$

with $\widehat{D} \in D$ since (\mathcal{L}, σ) is covered. Thus $C \sqsubseteq D$. Since $C \in \mathcal{C}$ was arbitrarily chosen, D is an upper bound of \mathcal{C} .

Now let E be any upper bound of \mathcal{C} . Consider an arbitrary $C \in \mathcal{C}$. There clearly is a $T \in \mathcal{T}$ with $C = [T]_s^\sigma$ and since E is an upper bound for \mathcal{C} there also are $T_C \in C$ and $T'_C \in E$ with $T_C \subseteq T'_C$.

We have $[T \cup T'_C]_s^\sigma = [T_C \cup T'_C]_s^\sigma = [T'_C]_s^\sigma = [E]_s^\sigma$ whence $T \cup T'_C \in E$. Now since every $C \in \mathcal{C}$ originates in some $T \in \mathcal{T}$ we get that for every $T \in \mathcal{T}$ there exists a $T'_C \in E$ with $T \cup T'_C \in E$, that is, $T \cup T'_C \subseteq \widehat{E}$. Then clearly $\bigcup_{T \in \mathcal{T}} T \subseteq \bigcup_{T \in \mathcal{T}} (T \cup T'_C) \subseteq \widehat{E}$, and by $\bigcup_{T \in \mathcal{T}} T \in D$ and $\widehat{E} \in E$ (the logic (\mathcal{L}, σ) is covered) we get $D \sqsubseteq E$. Since E was an arbitrarily chosen upper bound of \mathcal{C} , D is the least upper bound of \mathcal{C} . \square

While it preserves arbitrary joins, the mapping from theories to their strong equivalence classes does not necessarily preserve meets. It is easy to see that for the meet operation in the complete lattice of strong equivalence classes we have $[\bigcap_{T \in \mathcal{T}} T]_s^\sigma \sqsubseteq \bigcap_{T \in \mathcal{T}} [T]_s^\sigma$, as

$$\bigcap_{T \in \mathcal{T}} T \subseteq \bigcap_{T \in \mathcal{T}} \widehat{[T]_s^\sigma} \in \left[\bigcap_{T \in \mathcal{T}} \widehat{[T]_s^\sigma} \right]_s^\sigma = \bigcap_{T \in \mathcal{T}} [T]_s^\sigma$$

The reverse relation does not hold, as is witnessed by the following small, finite logic.

Example 2.35. Let $\mathcal{L} = \{a, b\}$ and \equiv_s^σ such that $\emptyset \not\equiv_s^\sigma \{a\} \equiv_s^\sigma \{b\} \equiv_s^\sigma \{a, b\}$. Then we get:

$$\begin{aligned} [\{a\} \cap \{b\}]_s^\sigma &= [\emptyset]_s^\sigma = \{\emptyset\} \\ [\{a\}]_s^\sigma \cap [\{b\}]_s^\sigma &= [\{a, b\}]_s^\sigma \cap [\{a, b\}]_s^\sigma = [\{a, b\}]_s^\sigma = \{\{a\}, \{b\}, \{a, b\}\} \neq \{\emptyset\} \end{aligned}$$

However, this is not a hindrance since the intersection property only needs to hold for arbitrary unions of theories and does not say anything about theory intersection.

We conclude this section with its main theorem showing that any full logic being covered possesses a characterization logic. The relevant characterization logic can even be defined (more or less) explicitly via a Herbrand-style canonical construction: roughly, the semantics of the characterization logic maps a theory T of the original language to the union of all strong equivalence classes $[S]_s^\sigma$ that are in \sqsubseteq -relation to the class $[T]_s^\sigma$ of the input theory.

Theorem 2.36. *Let (\mathcal{L}, σ) be a logic. If (\mathcal{L}, σ) is covered then a characterization logic for (\mathcal{L}, σ) is given by (\mathcal{L}, σ') with*

$$\sigma' : 2^{\mathcal{L}} \rightarrow 2^{2^{\mathcal{L}}}, \quad T \mapsto \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [T]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma$$

Proof. characterization: We have to show $\forall T_1, T_2 \subseteq \mathcal{L} : \sigma'(T_1) = \sigma'(T_2) \iff [T_1]_s^\sigma = [T_2]_s^\sigma$.

Let $T_1, T_2 \subseteq \mathcal{L}$.

\implies : Let $\sigma'(T_1) = \sigma'(T_2)$. Then by definition of σ' we get

$$\bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [T_1]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma = \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [T_2]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma$$

that is, for any $S \in 2^{\mathcal{L}}$ we find that $[T_1]_s^\sigma \subseteq [S]_s^\sigma$ iff $[T_2]_s^\sigma \subseteq [S]_s^\sigma$. The set $\{[S]_s^\sigma \mid S \in 2^{\mathcal{L}}\}$ corresponds to a partition of $2^{\mathcal{L}}$; thus for any element $T \in \bigcup_{S \in 2^{\mathcal{L}}, [T_1]_s^\sigma \subseteq [S]_s^\sigma} [S]_s^\sigma$ we have $[T_1]_s^\sigma \subseteq [T]_s^\sigma$, and if also $T \in \bigcup_{S \in 2^{\mathcal{L}}, [T_2]_s^\sigma \subseteq [S]_s^\sigma} [S]_s^\sigma$, then $[T_2]_s^\sigma \subseteq [T]_s^\sigma$ as well. Symmetry applies to obtain the above conclusion.

In particular, $T_1, T_2 \in 2^{\mathcal{L}}$ whence $[T_1]_s^\sigma \subseteq [T_1]_s^\sigma$ iff $[T_2]_s^\sigma \subseteq [T_1]_s^\sigma$, and $[T_1]_s^\sigma \subseteq [T_2]_s^\sigma$ iff $[T_2]_s^\sigma \subseteq [T_2]_s^\sigma$. This shows $[T_1]_s^\sigma \subseteq [T_2]_s^\sigma$ and $[T_2]_s^\sigma \subseteq [T_1]_s^\sigma$, that is, $[T_1]_s^\sigma = [T_2]_s^\sigma$.

\impliedby : Let $[T_1]_s^\sigma = [T_2]_s^\sigma$. Then immediately

$$\sigma'(T_1) = \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [T_1]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma = \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [T_2]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma = \sigma'(T_2)$$

intersection: We have to show $\forall \mathcal{T} \subseteq 2^{\mathcal{L}} : \sigma'(\bigcup_{T \in \mathcal{T}} T) = \bigcap_{T \in \mathcal{T}} \sigma'(T)$.

Let $\mathcal{T} \subseteq 2^{\mathcal{L}}$. From Theorem 2.34, we know that $[\bigcup_{T \in \mathcal{T}} T]_s^\sigma = \bigsqcup_{T \in \mathcal{T}} [T]_s^\sigma$ is the least upper bound of $\{[T]_s^\sigma \mid T \in \mathcal{T}\}$, whence for all $S \in 2^{\mathcal{L}}$, we have that $\bigsqcup_{T \in \mathcal{T}} [T]_s^\sigma \subseteq [S]_s^\sigma$ if and only if $[S]_s^\sigma$ is an upper bound of $\{[T]_s^\sigma \mid T \in \mathcal{T}\}$, that is,

$$\left[\bigcup_{T \in \mathcal{T}} T \right]_s^\sigma \subseteq [S]_s^\sigma \iff \forall T \in \mathcal{T} : [T]_s^\sigma \subseteq [S]_s^\sigma$$

Using this relationship, we get that

$$\begin{aligned} \sigma'\left(\bigcup_{T \in \mathcal{T}} T\right) &= \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [\bigcup_{T \in \mathcal{T}} T]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma \\ &= \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ \forall T \in \mathcal{T} : [T]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma \\ &= \bigcap_{T \in \mathcal{T}} \left(\bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [T]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma \right) \\ &= \bigcap_{T \in \mathcal{T}} \sigma'(T) \end{aligned}$$

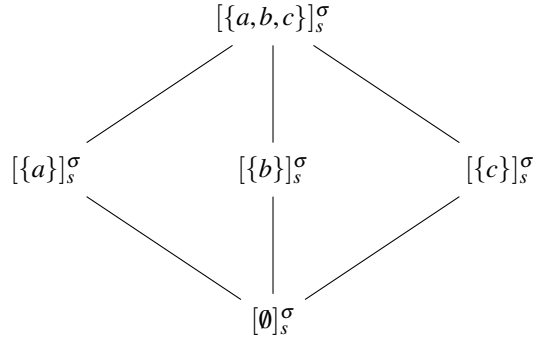
□

The construction from the statement of the theorem looks quite abstract, so we illustrate with a small concrete logic.

Example 2.37. We reconsider the logic from Example 2.26. There, $\mathcal{L} = \{a, b, c\}$ and

$$\begin{aligned} [\emptyset]_s^\sigma &= \{\emptyset\}, \\ [\{a\}]_s^\sigma &= \{\{a\}\}, \\ [\{b\}]_s^\sigma &= \{\{b\}\}, \\ [\{c\}]_s^\sigma &= \{\{c\}\}, \\ [\{a, b\}]_s^\sigma &= [\{a, c\}]_s^\sigma = [\{b, c\}]_s^\sigma = [\{a, b, c\}]_s^\sigma = \{\{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}. \end{aligned}$$

The resulting lattice of strong equivalence classes is depicted below:



According to Theorem 2.36, the characterization logic $\sigma' : 2^{\mathcal{L}} \rightarrow 2^{2^{\mathcal{L}}}$ assigns as follows:

$$\begin{aligned} \sigma'(\emptyset) &= \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [\emptyset]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma = \bigcup_{S \in 2^{\mathcal{L}}} [S]_s^\sigma = 2^{\mathcal{L}} \\ \sigma'(\{a\}) &= \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [\{a\}]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma = [\{a\}]_s^\sigma \cup [\{a, b, c\}]_s^\sigma = \{\{a\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\} \\ \sigma'(\{b\}) &= \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [\{b\}]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma = [\{b\}]_s^\sigma \cup [\{a, b, c\}]_s^\sigma = \{\{b\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\} \\ \sigma'(\{c\}) &= \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [\{c\}]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma = [\{c\}]_s^\sigma \cup [\{a, b, c\}]_s^\sigma = \{\{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\} \\ \sigma'(\{a, b, c\}) &= \bigcup_{\substack{S \in 2^{\mathcal{L}}, \\ [\{a, b, c\}]_s^\sigma \subseteq [S]_s^\sigma}} [S]_s^\sigma = [\{a, b, c\}]_s^\sigma = \{\{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\} \end{aligned}$$

Now it holds for example that

$$\begin{aligned}
 & \sigma'(\{a\}) \cap \sigma'(\{b\}) \\
 &= \{\{a\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\} \cap \{\{b\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\} \\
 &= \{\{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\} \\
 &= \sigma'(\{a,b\}) \\
 &= \sigma'(\{a\} \cup \{b\})
 \end{aligned}$$

2.2.2 Finite-Theory Characterization Logics

In the field of knowledge representation it is a common assumption that knowledge bases are finite. This is indeed not overly limiting, as finite knowledge bases will be most relevant for practical purposes. The following definition translates this assumption into our setting: the *finite-theory version* of a given logic (or simply, a *finite-theory logic*) considers only the finite knowledge bases of a language.

Definition 2.38. Given a full logic (\mathcal{L}, σ) , the finite-theory version $(\mathcal{L}, \sigma_{\text{fin}})$ of (\mathcal{L}, σ) is defined by the semantics

$$\sigma_{\text{fin}} : (2^{\mathcal{L}})_{\text{fin}} \rightarrow \sigma(2^{\mathcal{L}}) \quad \text{with} \quad \sigma_{\text{fin}}(T) = \sigma(T)$$

where $(2^{\mathcal{L}})_{\text{fin}} = \{T \in 2^{\mathcal{L}} \mid T \text{ is finite}\}$.

For finite-theory restrictions of logics, we adequately relax our requirements on characterization logics.

Definition 2.39. Let (\mathcal{L}, σ) be a full logic and $(\mathcal{L}, \sigma_{\text{fin}})$ its finite-theory version. We say that $(\mathcal{L}, \sigma'_{\text{fin}})$ is a *finite-theory characterization logic* for (\mathcal{L}, σ) if and only if:

1. $\forall T_1, T_2 \in (2^{\mathcal{L}})_{\text{fin}} : \sigma'_{\text{fin}}(T_1) = \sigma'_{\text{fin}}(T_2) \text{ iff } [T_1]_{\sigma'_{\text{fin}}} = [T_2]_{\sigma'_{\text{fin}}}$; (finite-theory characterization)
2. $\forall T_1, T_2 \in (2^{\mathcal{L}})_{\text{fin}} : \sigma'_{\text{fin}}(T_1 \cup T_2) = \sigma'_{\text{fin}}(T_1) \cap \sigma'_{\text{fin}}(T_2)$. (finite-theory intersection)

The second item requires binary intersection only; this is due to the fact that arbitrary unions of (finite) theories are not necessarily finite, and thus their semantics might not be well-defined.

As we did in the general case before, we first analyze the algebraic structure of the resulting model theories. We show that the model theory of any finite-theory characterization logic forms a lattice, that is, a partially ordered set where each non-empty finite subset has both a greatest lower bound and a least upper bound. (This is in contrast to *complete* lattices in the general case.) The proof is, although similar in procedure, slightly more involved than in the general case.

Proposition 2.40. Let $(\mathcal{L}, \sigma_{\text{fin}})$ be a finite-theory logic with characterization logic $(\mathcal{L}, \sigma'_{\text{fin}})$. Denoting $\mathcal{K} = \{\sigma'_{\text{fin}}(T) \mid T \in (2^{\mathcal{L}})_{\text{fin}}\}$, the pair (\mathcal{K}, \subseteq) is a lattice where glb and lub are given such that for all $K_1, K_2 \in \mathcal{K}$:

$$K_1 \wedge K_2 = K_1 \cap K_2 \text{ and } K_1 \vee K_2 = \bigwedge \{K_1, K_2\}^u$$

where $\{K_1, K_2\}^u = \{K \in \mathcal{K} \mid K_1 \subseteq K, K_2 \subseteq K\}$.

Proof. Let $K_1, K_2 \in \mathcal{K}$. Clearly there exist finite $T_1, T_2 \subseteq \mathcal{L}$ with $\sigma'_{\text{fin}}(T_1) = K_1$ as well as $\sigma'_{\text{fin}}(T_2) = K_2$.

glb: Obviously $T_1 \cup T_2 \in (2^{\mathcal{L}})_{\text{fin}}$. Therefore, it follows that

$$K_1 \wedge K_2 = \sigma'_{\text{fin}}(T_1) \cap \sigma'_{\text{fin}}(T_2) = \sigma'_{\text{fin}}(T_1) \cap \sigma'_{\text{fin}}(T_2) = \sigma'_{\text{fin}}(T_1 \cup T_2) \in \mathcal{K}$$

lub: It follows from the finite intersection property that σ'_{fin} is antimonotone. Now $\emptyset \subseteq T_1$ implies that $K_1 = \sigma'_{\text{fin}}(T_1) \subseteq \sigma'_{\text{fin}}(\emptyset)$ and likewise for T_2 . Thus $\sigma'_{\text{fin}}(\emptyset)$ is an upper bound of K_1 and K_2 , whence the set $\{K_1, K_2\}''$ is non-empty. We will now show that $\{K_1, K_2\}''$ is finite. Clearly both T_1 and T_2 have only finitely many subsets T'_1 and T'_2 . For each of these subsets, σ'_{fin} being antimonotone means that $\emptyset \subseteq T'_1 \subseteq T_1$ implies $\sigma'_{\text{fin}}(T_1) \subseteq \sigma'_{\text{fin}}(T'_1) \subseteq \sigma'_{\text{fin}}(\emptyset)$. Thus both K_1 and K_2 have only finitely many supersets in \mathcal{K} , that is, the sets K_1^\uparrow and K_2^\uparrow are finite, whence $\{K_1, K_2\}'' \subseteq K_1^\uparrow \cup K_2^\uparrow$ is finite. Consequently, $\{K_1, K_2\}''$ is a finite, non-empty subset of \mathcal{K} . It therefore possesses a greatest lower bound $K_l = \bigwedge \{K_1, K_2\}'' \in \mathcal{K}$. Since $\{K_1, K_2\}''$ is closed under intersection we have that $K_l = \bigwedge \{K_1, K_2\}'' = \bigcap \{K_1, K_2\}'' \in \{K_1, K_2\}''$ is the least element of $\{K_1, K_2\}''$ and therefore the least upper bound of K_1 and K_2 concluding the proof. \square

As before, we can show (with reasonable effort) that finite-theory characterization logics are unique up to isomorphism.

Theorem 2.41. *Let (\mathcal{L}, σ) be a finite-theory logic having two finite-theory characterization logics $(\mathcal{L}, \sigma'_{\text{fin}})$ and $(\mathcal{L}, \sigma''_{\text{fin}})$. Denoting the two carrier sets by $\mathcal{K}' = \{\sigma'_{\text{fin}}(T) \mid T \in (2^{\mathcal{L}})_{\text{fin}}\}$ and accordingly $\mathcal{K}'' = \{\sigma''_{\text{fin}}(T) \mid T \in (2^{\mathcal{L}})_{\text{fin}}\}$, the lattices $(\mathcal{K}', \subseteq)$ and $(\mathcal{K}'', \subseteq)$ are isomorphic.*

Proof. We provide a bijection $\phi : \mathcal{K}' \rightarrow \mathcal{K}''$ such that for all $K_1, K_2 \in \mathcal{K}'$, we find that $\phi(K_1 \wedge K_2) = \phi(K_1) \wedge \phi(K_2)$ and $\phi(K_1 \vee K_2) = \phi(K_1) \vee \phi(K_2)$. Let $K \in \mathcal{K}'$. By definition, there exists a finite $T \subseteq \mathcal{L}$ with $\sigma'_{\text{fin}}(T) = K$. Define $\phi(K) = \sigma''_{\text{fin}}(T)$.

ϕ is bijective: The proof is as in the general case.

ϕ is structure-preserving: Let $K_1, K_2 \in \mathcal{K}'$. Clearly there exist $T_1, T_2 \subseteq \mathcal{L}$ such that $\sigma'_{\text{fin}}(T_1) = K_1$ and $\sigma'_{\text{fin}}(T_2) = K_2$. We have

$$\begin{aligned}
 \phi(K_1 \wedge K_2) &= \phi(K_1 \cap K_2) && \text{(Def. } \wedge) \\
 &= \phi(\sigma'_{\text{fin}}(T_1) \cap \sigma'_{\text{fin}}(T_2)) && (\sigma'_{\text{fin}} \text{ is onto } \mathcal{K}') \\
 &= \phi(\sigma'_{\text{fin}}(T_1 \cup T_2)) && (\text{intersection } \sigma'_{\text{fin}}) \\
 &= \sigma''_{\text{fin}}(T_1 \cup T_2) && \text{(Def. } \phi) \\
 &= \sigma''_{\text{fin}}(T_1) \cap \sigma''_{\text{fin}}(T_2) && (\text{intersection } \sigma''_{\text{fin}}) \\
 &= \phi(\sigma'_{\text{fin}}(T_1)) \cap \phi(\sigma'_{\text{fin}}(T_2)) && \text{(Def. } \phi) \\
 &= \phi(K_1) \cap \phi(K_2) && (\text{assumption}) \\
 &= \phi(K_1) \wedge \phi(K_2) && \text{(Def. } \wedge)
 \end{aligned}$$

With reasonable effort, we can also show that

$$\begin{aligned}
 & \phi(K_1 \vee K_2) \\
 &= \phi\left(\bigwedge \{K_1, K_2\}^u\right) \\
 &\quad (\text{Def. } \vee) \\
 &= \bigwedge \phi(\{K_1, K_2\}^u) \\
 &\quad (\phi \text{ preserves } \wedge) \\
 &= \bigwedge \phi(\{K \in \mathcal{K}' \mid K_1 \subseteq K, K_2 \subseteq K\}) \\
 &\quad (\text{Def. } \cdot^u) \\
 &= \bigwedge \{\phi(K) \mid K \in \mathcal{K}', K_1 \subseteq K, K_2 \subseteq K\} \\
 &\quad (\text{notation}) \\
 &= \bigwedge \left\{ \phi(\sigma'_{\text{fin}}(T)) \mid T \in \left(2^{\mathcal{L}}\right)_{\text{fin}}, \sigma'_{\text{fin}}(T_1) \subseteq \sigma'_{\text{fin}}(T), \sigma'_{\text{fin}}(T_2) \subseteq \sigma'_{\text{fin}}(T) \right\} \\
 &\quad (\sigma'_{\text{fin}} \text{ is onto } \mathcal{K}') \\
 &= \bigwedge \left\{ \phi(\sigma'_{\text{fin}}(T)) \mid T \in \left(2^{\mathcal{L}}\right)_{\text{fin}}, \sigma'_{\text{fin}}(T_1) \cap \sigma'_{\text{fin}}(T) = \sigma'_{\text{fin}}(T_1), \sigma'_{\text{fin}}(T_2) \cap \sigma'_{\text{fin}}(T) = \sigma'_{\text{fin}}(T_2) \right\} \\
 &\quad (\text{elementary}) \\
 &= \bigwedge \left\{ \phi(\sigma'_{\text{fin}}(T)) \mid T \in \left(2^{\mathcal{L}}\right)_{\text{fin}}, \sigma'_{\text{fin}}(T_1 \cup T) = \sigma'_{\text{fin}}(T_1), \sigma'_{\text{fin}}(T_2 \cup T) = \sigma'_{\text{fin}}(T_2) \right\} \\
 &\quad (\text{intersection } \sigma'_{\text{fin}}) \\
 &= \bigwedge \left\{ \phi(\sigma'_{\text{fin}}(T)) \mid T \in \left(2^{\mathcal{L}}\right)_{\text{fin}}, [T_1 \cup T]_s^\sigma = [T_1]_s^\sigma, [T_2 \cup T]_s^\sigma = [T_2]_s^\sigma \right\} \\
 &\quad (\text{characterization } \sigma'_{\text{fin}}) \\
 &= \bigwedge \left\{ \phi(\sigma'_{\text{fin}}(T)) \mid T \in \left(2^{\mathcal{L}}\right)_{\text{fin}}, \sigma''_{\text{fin}}(T_1 \cup T) = \sigma''_{\text{fin}}(T_1), \sigma''_{\text{fin}}(T_2 \cup T) = \sigma''_{\text{fin}}(T_2) \right\} \\
 &\quad (\text{characterization } \sigma''_{\text{fin}}) \\
 &= \bigwedge \left\{ \sigma''_{\text{fin}}(T) \mid T \in \left(2^{\mathcal{L}}\right)_{\text{fin}}, \sigma''_{\text{fin}}(T_1) \cap \sigma''_{\text{fin}}(T) = \sigma''_{\text{fin}}(T_1), \sigma''_{\text{fin}}(T_2) \cap \sigma''_{\text{fin}}(T) = \sigma''_{\text{fin}}(T_2) \right\} \\
 &\quad (\text{intersection } \sigma''_{\text{fin}}) \\
 &= \bigwedge \left\{ \sigma''_{\text{fin}}(T) \mid T \in \left(2^{\mathcal{L}}\right)_{\text{fin}}, \sigma''_{\text{fin}}(T_1) \subseteq \sigma''_{\text{fin}}(T), \sigma''_{\text{fin}}(T_2) \subseteq \sigma''_{\text{fin}}(T) \right\} \\
 &\quad (\text{elementary}) \\
 &= \bigwedge \{K \mid K \in \mathcal{K}'', \sigma''_{\text{fin}}(T_1) \subseteq K, \sigma''_{\text{fin}}(T_2) \subseteq K\} \\
 &\quad (\sigma''_{\text{fin}} \text{ is onto } \mathcal{K}'') \\
 &= \bigwedge \{\sigma''_{\text{fin}}(T_1), \sigma''_{\text{fin}}(T_2)\}^u \\
 &\quad (\text{Def. } \cdot^u) \\
 &= \sigma''_{\text{fin}}(T_1) \vee \sigma''_{\text{fin}}(T_2) \\
 &\quad (\text{Def. } \vee) \\
 &= \phi(\sigma'_{\text{fin}}(T_1)) \vee \phi(\sigma'_{\text{fin}}(T_2)) \\
 &\quad (\text{Def. } \phi) \\
 &= \phi(K_1) \vee \phi(K_2) \\
 &\quad (\text{assumption})
 \end{aligned}$$

Thus ϕ is a structure-preserving bijection from \mathcal{K}' to \mathcal{K}'' , and the two lattices are isomorphic. \square

The following theorem shows that any logic possesses a finite-theory characterization logic. This means that the most important case for knowledge representation behaves well in the sense that characterization logics always exist.

Theorem 2.42. *Let (\mathcal{L}, σ) be a full logic. Then a finite-theory characterization logic for (\mathcal{L}, σ) is given by $(\mathcal{L}, \sigma'_{fin})$ with*

$$\sigma'_{fin} : (2^{\mathcal{L}})_{fin} \rightarrow 2^{2^{\mathcal{L}}}, \quad T \mapsto \bigcup_{\substack{S \in (2^{\mathcal{L}})_{fin}, \\ T \subseteq S}} [S]_s^{\sigma_{fin}}$$

Proof. finite intersection: We show that for all $T \in (2^{\mathcal{L}})_{fin}$, we find $\sigma'_{fin}(T) = \bigcap_{t \in T} \sigma'_{fin}(\{t\})$. Consider $T \in (2^{\mathcal{L}})_{fin}$.

\subseteq : Let $U \in \sigma'_{fin}(T)$. Hence, there is an $S \in (2^{\mathcal{L}})_{fin}$ such that $U \in [S]_s^{\sigma_{fin}}$ and $T \subseteq S$. Consequently, for any $t \in T$, we find $\{t\} \subseteq T \subseteq S$ and thus $[S]_s^{\sigma_{fin}} \subseteq \sigma'_{fin}(\{t\})$ showing $U \in \bigcap_{t \in T} \sigma'_{fin}(\{t\})$.

\supseteq : Consider now $U \in \bigcap_{t \in T} \sigma'_{fin}(\{t\})$. This means that for any $t \in T$ there exists an $S_t \in (2^{\mathcal{L}})_{fin}$ such that $U \in [S_t]_s^{\sigma_{fin}}$ (that is, $[U]_s^{\sigma_{fin}} = [S_t]_s^{\sigma_{fin}}$) and $\{t\} \subseteq S_t$. It follows that for any $t, t' \in T$, we find $[S_t]_s^{\sigma_{fin}} = [U]_s^{\sigma_{fin}} = [S_{t'}]_s^{\sigma_{fin}}$. Let us fix a certain $\bar{t} \in T$. In particular, $U \in [S_{\bar{t}}]_s^{\sigma_{fin}}$. Now consider the set $\bigcup_{t \in T} S_t$, which is finite since T is finite and each S_t is finite. Furthermore, $\bigcup_{t \in T} S_t \in [S_{\bar{t}}]_s^{\sigma_{fin}}$, that is, $[\bigcup_{t \in T} S_t]_s^{\sigma_{fin}} = [S_{\bar{t}}]_s^{\sigma_{fin}}$ by Theorem 2.29. Since for all $t \in T$ we have $t \in S_t$, we conclude that $T = \bigcup_{t \in T} \{t\} \subseteq \bigcup_{t \in T} S_t$. In turn, this yields $U \in \sigma'_{fin}(T)$.

finite-theory characterization: We show that for all $T_1, T_2 \in (2^{\mathcal{L}})_{fin}$, we find $[T_1]_s^{\sigma_{fin}} = [T_2]_s^{\sigma_{fin}}$ if and only if $\sigma'_{fin}(T_1) = \sigma'_{fin}(T_2)$.

if: Let $T_1, T_2 \in (2^{\mathcal{L}})_{fin}$ with $\sigma'_{fin}(T_1) = \sigma'_{fin}(T_2)$. Firstly, the definition of σ'_{fin} yields that for each $S \in (2^{\mathcal{L}})_{fin}$ we have $[S]_s^{\sigma_{fin}} \subseteq \sigma'_{fin}(S)$. In combination with the presumption, this means that $[T_1]_s^{\sigma_{fin}} \subseteq \sigma'_{fin}(T_1) = \sigma'_{fin}(T_2)$ and $[T_2]_s^{\sigma_{fin}} \subseteq \sigma'_{fin}(T_2) = \sigma'_{fin}(T_1)$. Hence, there is a set $S_1 \in [T_2]_s^{\sigma_{fin}}$ with $T_1 \subseteq S_1$, whence $[T_1]_s^{\sigma_{fin}} \subseteq [S_1]_s^{\sigma_{fin}} = [T_2]_s^{\sigma_{fin}}$. Likewise, there is a set $S_2 \in [T_1]_s^{\sigma_{fin}}$ with $T_2 \subseteq S_2$, whence also $[T_2]_s^{\sigma_{fin}} \subseteq [S_2]_s^{\sigma_{fin}} = [T_1]_s^{\sigma_{fin}}$. By Theorem 2.31 (saying that \subseteq is antisymmetric), we get $[T_1]_s^{\sigma_{fin}} = [T_2]_s^{\sigma_{fin}}$.

only if: Let $T_1, T_2 \in (2^{\mathcal{L}})_{fin}$ with $[T_1]_s^{\sigma_{fin}} = [T_2]_s^{\sigma_{fin}}$ and consider any $T \in \sigma'_{fin}(T_1)$. Then there is an $S \in (2^{\mathcal{L}})_{fin}$ such that $T \in [S]_s^{\sigma_{fin}}$ and $T_1 \subseteq S$. Clearly there is a $U \in (2^{\mathcal{L}})_{fin}$ with $T_1 \cup U = S$. Since $[T_1]_s^{\sigma_{fin}} = [T_2]_s^{\sigma_{fin}}$ by presumption, we can apply Lemma 2.29, which yields $T \in [S]_s^{\sigma_{fin}} = [T_1 \cup U]_s^{\sigma_{fin}} = [T_2 \cup U]_s^{\sigma_{fin}}$. Obviously, we have $T_2 \cup U \in (2^{\mathcal{L}})_{fin}$ with $T_2 \subseteq T_2 \cup U$ and the definition of σ'_{fin} yields $T \in \sigma'_{fin}(T_2)$. This shows $\sigma'_{fin}(T_1) \subseteq \sigma'_{fin}(T_2)$; the reverse inclusion $\sigma'_{fin}(T_2) \subseteq \sigma'_{fin}(T_1)$ holds by symmetry. \square

Intuitively, in this canonical construction of a characterization semantics σ'_{fin} (akin to Herbrand interpretations in first-order logic), the model set of a theory T is the set of all theories that are strongly equivalent to some supertheory of T .

2.3 Applying Canonical Constructions to Nonmonotonic Formalisms

In the previous section we have seen that (under certain condition) the existence of characterization logics for knowledge representation formalisms are guaranteed. These results are achieved by defining in a sense Herbrand-style canonical construction. More precisely, the characterization semantics of the new logic is defined in terms of certain unions of strong equivalence classes of the original language. Such a characterization semantics is usually far from being intuitive or self-explanatory. The intended role of this semantics was to serve as a witness for the existence of characterization logics. Nevertheless, we will discuss the application of our general, abstract results to some of the formalisms presented in Example 2.1. We start with abstract argumentation theory, which is a vibrant as well as immensely growing research area in AI [Dung, 1995]. Surprisingly, we get a meaningful result very similar to the recently introduced Dung logics [Baumann and Brewka, 2015].

2.3.1 Abstract Argumentation Theory

We start with a very brief introduction to Dung's argumentation theory which is sufficient for the moment (cf. Section 3.1 for more detailed as well as more general definitions and [Baroni et al., 2018a] for a recent and comprehensive overview).

An *argumentation framework* (AF) is a pair $F = (A, R)$ such that $R \subseteq A \times A$. Although there exists some work on unrestricted AFs [Baumann and Spanring, 2015; Baumann and Spanring, 2017] it is common to assume that A , the set of arguments, is a finite subset of a fixed infinite background set \mathcal{U} . Let us denote the class of all finite AFs by AF_{fin} . An (extension-based) *argumentation semantics* is a function $\rho : AF_{\text{fin}} \rightarrow 2^{\mathcal{U}}$ where elements of $\rho(F)$ are called ρ -*extensions* of F . The most prominent one is stable semantics (abbreviated by *stb*) which was already defined by Dung in 1995. A set E is a *stb*-extension of F if 1. there are no $a, b \in E$, s.t. $(a, b) \in R$ (*conflict-freeness*) and 2. for any $c \in A \setminus E$, there is an $a \in E$, s.t. $(a, c) \in R$ (*full range*).

We proceed with some notational conventions and the precise definition of strong equivalence in case of AFs. The union $F \cup G$ as well as subset-relation $F \subseteq G$ of two AFs is understood to be pointwise, that is, $(A_1, R_1) \cup (A_2, R_2) = (A_1 \cup A_2, R_1 \cup R_2)$, and, similarly, $(A_1, R_1) \subseteq (A_2, R_2)$ if and only if $A_1 \subseteq A_2$ and $R_1 \subseteq R_2$.

Definition 2.43. Given an argumentation semantics ρ . Two AFs F and G are *strongly ρ -equivalent* if for any $H \in AF_{\text{fin}}$, $\rho(F \cup H) = \rho(G \cup H)$. For short, $F \equiv_s^\rho G$.

The first work regarding characterizing strong equivalence for AFs was [Oikarinen and Woltran, 2011]. It turned out that deciding this notion is deeply linked to the syntax of AFs. In general, any argument being part of an AF may contribute towards future extensions. However, for each semantics, there are patterns of redundant attacks captured by so-called *kernels*. Formally, a kernel is a function $k : AF_{\text{fin}} \rightarrow AF_{\text{fin}}$ where $k(F) = F^k$ is obtained from F by deleting certain redundant attacks. Consider the following definition and characterization theorem.

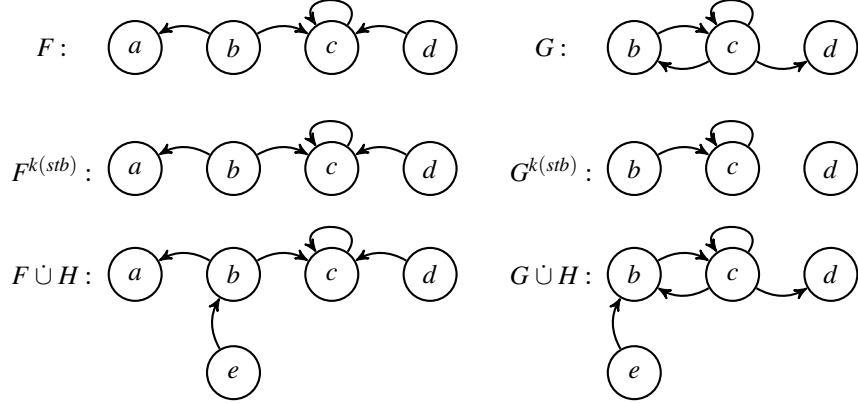
Definition 2.44. Given an AF $F = (A, R)$. The *stb-kernel* $F^{k(\text{stb})} = (A, R^{k(\text{stb})})$ is defined with $R^{k(\text{stb})} = R \setminus \{(a, b) \mid a \neq b \wedge (a, a) \in R\}$.

Theorem 2.45 ([Oikarinen and Woltran, 2011]). *For two AFs F, G we have:*

$$F \equiv_s^{\text{stb}} G \iff F^{k(\text{stb})} = G^{k(\text{stb})}.$$

Let us illustrate the introduced concepts with an example.

Example 2.46. Consider the following six AFs.



The AFs F and G possess the same stable extensions, namely $stb(F) = stb(G) = \{\{b, d\}\}$. However, both frameworks are not strongly stb -equivalent since $F^{k(stb)} \neq G^{k(stb)}$. A witnessing framework is given by H . Indeed $stb(F \cup H) = \{\{a, d, e\}\} \neq \emptyset = stb(G \cup H)$.

Let us consider now argumentation theory in the general setup. In Example 2.1 we have seen that the embedding of abstract argumentation in our setting is a bit more involved than in case of propositional logic, logic programs or default logic. The main reason for this is that in contrast to the other considered formalisms we have that abstract argumentation frameworks possess two sorts of building blocks, namely arguments and attacks. Moreover, the latter are dependent since adding attacks requires the presence of the corresponding arguments. In order to cast AFs into our general setup we have to have theories which correspond to AFs, s.t. the standard set union \cup of such theories correspond to $\dot{\cup}$ on the AF-level. Moreover, the semantics of theories has to correspond to the argumentation semantics of the associated AFs.

We start with the introduction of a ρ -logic which formally captures a specific argumentation semantics ρ on the level of theories.

Definition 2.47. Let \mathcal{U} be a background set of arguments and ρ be an AF semantics. A ρ -logic is a triple $(\mathcal{L}_{AF}, I, \sigma_\rho)$ where $\mathcal{L}_{AF} = \{(\{a\}, \emptyset), (\{a, b\}, \{(a, b)\}) \mid a, b \in \mathcal{U}\}$, $I = 2^{\mathcal{U}}$ and $\sigma_\rho : 2^{\mathcal{L}_{AF}} \rightarrow 2^I$ with $\sigma_\rho(T) = \rho(\bigcup_{t \in T} t)$.

The representational issue implies that two different theories may represent the same framework which causes some additional effort. More precisely, in what follows it will be a typical task to show that the presented results are independent of the concrete representation of a certain framework. Let us start with an illustrating example.

Example 2.48. Let $T = \{(\{b, c\}, \{(b, c)\}), (\{c, b\}, \{(c, b)\}), (\{c, d\}, \{(c, d)\}), (\{c\}, \{(c, c)\})\}$ and $S = T \cup \{(\{a\}, \emptyset), (\{b\}, \emptyset), (\{c\}, \emptyset)\}$. Please observe that $\bigcup_{t \in T} t = \bigcup_{s \in S} s = G$ as depicted in Example 2.46. Moreover by definition we have $\sigma_{stb}(T) = \sigma_{stb}(S) = stb(G) = \{\{b, d\}\}$.

The following functions (restricted to the finite case) will be frequently used. First, we define the *associated AF* of a given theory via $AF : (2^{\mathcal{L}_{AF}})_{\text{fin}} \rightarrow AF_{\text{fin}}$ where $T \mapsto \bigcup_{t \in T} t$. As already discussed the function $AF(\cdot)$ is not injective as demonstrated in Example 2.48. Secondly, the *canonical representation* of a given AF is defined by $C : AF_{\text{fin}} \rightarrow (2^{\mathcal{L}_{AF}})_{\text{fin}}$ where (A, R) is represented by the \mathcal{L}_{AF} -theory $\{(\{a\}, \emptyset) \mid a \in A\} \cup \{(\{a, b\}, \{(a, b)\}) \mid (a, b) \in R\}$. Observe that for any AF H , we find $AF(C(H)) = H$. Moreover, regarding Examples 2.46 and 2.48 we have $C(G) = S \neq T$.

Note that the assumption of finiteness of AFs can be reflected by considering the finite-theory versions of ρ -logics. Before applying our canonical construction presented in Theorem 2.42 we have to ensure that a constructed ρ -logic correctly reflects AFs under semantics ρ . We start with two simple properties showing that the concrete representation (of an AF via a theory) is not “seen” by set union as well as semantics σ_ρ .

Proposition 2.49. *Let $(\mathcal{L}_{AF}, I, \sigma_\rho)$ be a ρ -logic and consider any theories $S, T \subseteq \mathcal{L}_{AF}$.*

1. $AF(S \cup T) = AF(S) \dot{\cup} AF(T)$ and
2. $\sigma_\rho(S \cup T) = \rho(AF(S) \dot{\cup} AF(T))$.

Proof. 1. Both statements can be easily seen. Consider the following equations.

$$\begin{aligned}
 & AF(S \cup T) \\
 &= \dot{\bigcup}_{u \in S \cup T} u && \text{(Definition } AF) \\
 &= \dot{\bigcup}_{s \in S} s \dot{\cup} \dot{\bigcup}_{t \in T} t && \text{(associativity } \dot{\bigcup}) \\
 &= AF(S) \dot{\cup} AF(T) && \text{(Definition } AF)
 \end{aligned}$$

2.

$$\begin{aligned}
 & \sigma_\rho(S \cup T) \\
 &= \rho\left(\dot{\bigcup}_{u \in S \cup T} u\right) && \text{(Definition } \sigma_\rho) \\
 &= \rho(AF(S \cup T)) && \text{(Definition } AF) \\
 &= \rho(AF(S) \dot{\cup} AF(T)) && \text{(Item 1)}
 \end{aligned}$$

□

The following theorem shows that two \mathcal{L}_{AF} -theories S and T are strongly equivalent under σ_ρ if and only if the AFs $AF(S)$ and $AF(T)$ are strongly equivalent under ρ (denoted by $AF(S) \equiv_s^\rho AF(T)$). We mention that the theorem does not require finiteness and is thus valid for arbitrary cardinalities of theories as well as AFs.

Theorem 2.50. *Let $(\mathcal{L}_{AF}, I, \sigma_\rho)$ be a ρ -logic. For $S, T \subseteq \mathcal{L}_{AF}$ we have*

$$AF(S) \equiv_s^\rho AF(T) \iff [S]_s^{\sigma_\rho} = [T]_s^{\sigma_\rho}.$$

Proof. \implies : Let $AF(S) \equiv_s^\rho AF(T)$ and $V \in 2^{\mathcal{L}_{AF}}$. We have to show that both theories are strongly equivalent, i.e. $\sigma_\rho(S \cup V) = \sigma_\rho(T \cup V)$.

$$\begin{aligned}
 & \sigma_\rho(S \cup V) \\
 &= \rho(AF(S) \dot{\cup} AF(V)) && \text{(Observation 2.49)} \\
 &= \rho(AF(T) \dot{\cup} AF(V)) && \text{(assumption } AF(S) \equiv_s^\rho AF(T)) \\
 &= \sigma_\rho(T \cup V) && \text{(Observation 2.49)}
 \end{aligned}$$

\Leftarrow : Let $[S]_s^{\sigma_p} = [T]_s^{\sigma_p}$ and H be an AF. We have to show that the corresponding AFs are strongly equivalent, i.e. $\rho(AF(S) \dot{\cup} H) = \rho(AF(T) \dot{\cup} H)$.

$$\begin{aligned}
 & \rho(AF(S) \dot{\cup} H) \\
 &= \rho(AF(S) \dot{\cup} AF(C(H))) && (AF(C(H)) = H) \\
 &= \sigma_p(S \cup C(H)) && (\text{Observation 2.49}) \\
 &= \sigma_p(T \cup C(H)) && (\text{assumption } [S]_s^{\sigma_p} = [T]_s^{\sigma_p}) \\
 &= \rho(AF(T) \dot{\cup} AF(C(H))) && (\text{Observation 2.49}) \\
 &= \rho(AF(T) \dot{\cup} H) && (AF(C(H)) = H)
 \end{aligned}$$

□

Due to Theorem 2.42 we are able to present finite-theory characterization logics for any ρ -logic.

Corollary 2.51. *Let $(\mathcal{L}_{AF}, I, \sigma_p)$ be a ρ -logic. The following logic $(\mathcal{L}_{AF}, \kappa)$ is a finite-theory characterization logic of $(\mathcal{L}_{AF}, I, \sigma_p)$:*

$$\kappa : \left(2^{\mathcal{L}_{AF}}\right)_{fin} \rightarrow 2^{2^{\mathcal{L}_{AF}}}, \quad T \mapsto \bigcup_{\substack{S \in \left(2^{\mathcal{L}_{AF}}\right)_{fin}, \\ T \subseteq S}} [S]_s^{(\sigma_p)_{fin}}$$

So far, so good, but how can we interpret these finite-theory characterization logics in terms of argumentation theory? In other words, what is the corresponding characterization semantics on the level of pure AFs (instead of theories associated with AFs)? We extend the function AF to sets of theories as usual, namely $AF : 2^{2^{\mathcal{L}_{AF}}} \rightarrow 2^{AF_{fin}}$ where $\mathcal{T} \mapsto \{AF(T) \mid T \in \mathcal{T}\}$.⁴ Consider the following definition. We will see that all crucial properties of κ transfer to ρ' , that is, ρ' satisfies finite intersection and furthermore, it characterizes strong equivalence under ρ .

Definition 2.52. Given an argumentation semantics $\rho : AF_{fin} \rightarrow 2^{\mathcal{U}}$. We define $\rho' : AF_{fin} \rightarrow 2^{AF_{fin}}$ with $F \mapsto AF(\kappa(C(F)))$.

Proposition 2.53. *For any argumentation semantics ρ and semantics ρ' as defined above we have:*

1. $\forall F, G \in AF_{fin} : \rho'(F) = \rho'(G) \iff F \equiv_s^\rho G;$ (finite-theory characterization)
2. $\forall F, G \in AF_{fin} : \rho'(F \dot{\cup} G) = \rho'(F) \cap \rho'(G).$ (finite-theory intersection)

Proof. finite-theory characterization: \Leftarrow : Let $F \equiv_s^\rho G$. Hence, $AF(C(F)) \equiv_s^\rho AF(C(G))$. Consequently, $[C(F)]_s^{\sigma_p} = [C(G)]_s^{\sigma_p}$ (Theorem 2.50). Since $(\mathcal{L}_{AF}, \kappa)$ is a finite-theory characterization logic of $(\mathcal{L}_{AF}, I, \sigma_p)$ (Definition 2.51) we deduce that $\kappa(C(F)) = \kappa(C(G))$. Obviously, $AF(\kappa(C(F))) = AF(\kappa(C(G)))$ which means $\rho'(F) = \rho'(G)$ (Definition 2.52).

⁴We do not introduce a new symbol for the new function. Which function is meant will be clear from the context.

\implies : We prove the contrapositive. Hence, let $F \not\equiv_s^p G$. This means, $AF(C(F)) \not\equiv_s^p AF(C(G))$ and we obtain $[C(F)]_s^{\sigma_p} \neq [C(G)]_s^{\sigma_p}$ (Theorem 2.50). Consequently, $\kappa(C(F)) \neq \kappa(C(G))$ since κ characterizes strong equivalence under σ_p (Definition 2.51). Since equivalence classes are disjoint we deduce the existence of a theory U , such that (without loss of generality) $[U]_s^{\sigma_p} \subseteq \kappa(C(F)) \setminus \kappa(C(G))$. Consequently, $AF(U) \in AF(\kappa(C(F))) \setminus AF(\kappa(C(G)))$ since for all other representations of U' , s.t. $AF(U) = AF(U')$, we have $U' \in [U]_s^{\sigma_p}$ (Theorem 2.50). Hence, $AF(\kappa(C(F))) \neq AF(\kappa(C(G)))$ which means $\rho'(F) \neq \rho'(G)$ (Definition 2.52).

finite-theory intersection:

$$\begin{aligned}
 & \rho'(F \dot{\cup} G) \\
 &= AF(\kappa(C(F \dot{\cup} G))) && \text{(Definition 2.52)} \\
 &= AF(\kappa(C(F) \cup C(G))) && (C(F \dot{\cup} G) = C(F) \cup C(G)) \\
 &= AF(\kappa(C(F)) \cap \kappa(C(G))) && \text{(intersection } \kappa) \\
 &= AF(\kappa(C(F))) \cap AF(\kappa(C(G))) && \text{(can be seen)} \\
 &= \rho'(F) \cap \rho'(G) && \text{(Definition 2.52)}
 \end{aligned}$$

□

Finally, we present an equivalent definition of ρ' that does not rely on ρ -logics. This means the evaluation of ρ' can be done purely on the level of AFs.

Proposition 2.54. *Let $\rho : AF_{fin} \rightarrow 2^{AF_{fin}}$ be a semantics and ρ' as in Definition 2.52. For any $F \in AF_{fin}$ we have:*

$$\rho'(F) = \bigcup_{\substack{G \in AF_{fin}, \\ F \dot{\subseteq} G}} \{H \mid H \equiv_s^p G\}$$

Proof.

$$\begin{aligned}
 \rho'(F) &= AF(\kappa(C(F))) && \text{(Definition 2.52)} \\
 &= AF\left(\bigcup_{\substack{S \in (2^{\mathcal{L}_{AF}})_{\text{fin}}, \\ C(F) \subseteq S}} [S]_s^{(\sigma_\rho)_{\text{fin}}}\right) && \text{(Definition 2.51)} \\
 &= AF\left(\bigcup_{\substack{S \in (2^{\mathcal{L}_{AF}})_{\text{fin}}, \\ C(F) \subseteq C(AF(S))}} [S]_s^{(\sigma_\rho)_{\text{fin}}}\right) && (S \subseteq C(AF(S)), [S]_s^{(\sigma_\rho)_{\text{fin}}} = [C(AF(S))]_s^{(\sigma_\rho)_{\text{fin}}}) \\
 &= \bigcup_{\substack{S \in (2^{\mathcal{L}_{AF}})_{\text{fin}}, \\ C(F) \subseteq C(AF(S))}} AF([S]_s^{(\sigma_\rho)_{\text{fin}}}) && \text{(Definition } AF(\cdot)\text{)} \\
 &= \bigcup_{\substack{S \in (2^{\mathcal{L}_{AF}})_{\text{fin}}, \\ F \dot{\subseteq} AF(S)}} AF([S]_s^{(\sigma_\rho)_{\text{fin}}}) && (C(F) \subseteq C(AF(S)) \iff F \dot{\subseteq} AF(S)) \\
 &= \bigcup_{\substack{S \in (2^{\mathcal{L}_{AF}})_{\text{fin}}, \\ F \dot{\subseteq} AF(S)}} \{AF(T) \mid T \in [S]_s^{(\sigma_\rho)_{\text{fin}}}\} && \text{(Definition } AF(\cdot)\text{)} \\
 &= \bigcup_{\substack{S \in (2^{\mathcal{L}_{AF}})_{\text{fin}}, \\ F \dot{\subseteq} AF(S)}} \{AF(T) \mid [T]_s^{(\sigma_\rho)_{\text{fin}}} = [S]_s^{(\sigma_\rho)_{\text{fin}}}\} && \text{(equivalence relation)} \\
 &= \bigcup_{\substack{G \in AF_{\text{fin}}, \\ F \dot{\subseteq} G}} \{H \mid H \equiv_s^p G\} && (AF(S) = G, AF(T) = H, \text{Theorem 2.50})
 \end{aligned}$$

□

Recently, so-called *Dung-logics* were introduced to be able to perform AGM-style revision for Dung's abstract argumentation frameworks [Baumann and Brewka, 2015]. These Dung-logics are very similar yet different from the characterization logics presented in Proposition 2.54. The main difference is that theories in Dung-logics are sets of AFs in contrast to the newly presented characterization logic where theories correspond to single AFs. We mention that Dung-logics possess the intersection property (Definition 2.55) and furthermore, two AFs F and G are strongly equivalent with respect to an argumentation semantics ρ if and only if the singletons of F and G are ordinarily equivalent with respect to the semantics defined by Theorem 3 of [Baumann and Brewka, 2015].

We start with the formal definition of a Dung-logic in case of the most prominent argumentation semantics, namely stable semantics.

Definition 2.55. The Dung-logic in case of stable semantics is a pair $(AF_{\text{fin}}, \delta)$ with

$$\delta : 2^{AF_{\text{fin}}} \rightarrow 2^{AF_{\text{fin}}} \quad \mathcal{U} \mapsto \bigcap_{F \in \mathcal{U}} \text{Mod}^{k(stb)}(F)$$

whereas $\text{Mod}^{k(stb)}(F) = \left\{ G \in AF_{\text{fin}} \mid F^{k(stb)} \subseteq G^{k(stb)} \right\}$.

In order to see the similarity we consider the above definition for singletons of AFs.

Observation 1. Let $stb : AF_{\text{fin}} \rightarrow 2^{AF_{\text{fin}}}$ be stable semantics and δ as in Definition 2.55. For any $F \in AF_{\text{fin}}$ we have:

$$\delta(\{F\}) = \bigcup_{\substack{G \in AF_{\text{fin}}, \\ F^{k(stb)} \subseteq G^{k(stb)}}} \{H \mid H \equiv_s^{stb} G\}$$

The only difference in comparison to ρ' is that for δ , we include all equivalence classes of AFs G whose *kernels* are in superset relation with the kernel of F , instead of having the superset relation on the AFs themselves. A further analysis will be part of future work.

2.3.2 Normal Logic Programs

For normal logic programs under the stable model semantics as presented in Example 2.1, applying Theorem 2.42 yields:

Corollary 2.56. For the finite-theory version of logic $(\mathcal{L}_{LP}, \sigma_{stb})$ of normal logic programs under stable model semantics, a finite-theory characterization logic is given by

$$\sigma'_{stb} : \left(2^{\mathcal{L}_{LP}}\right)_{\text{fin}} \rightarrow 2^{\left(2^{\mathcal{L}_{LP}}\right)_{\text{fin}}}, \quad T \mapsto \bigcup_{\substack{S \in \left(2^{\mathcal{L}_{LP}}\right)_{\text{fin}}, \\ T \subseteq S}} [S]_s^{\sigma_{stb}}$$

An existing, well-known characterizing semantics is given by SE-models.

Definition 2.57. Let P be a normal logic program over \mathcal{A} and $X \subseteq Y \subseteq A$. Define semantics $\sigma_{SE} : \left(2^{\mathcal{L}_{LP}}\right)_{\text{fin}} \rightarrow 2^{\mathcal{A} \times \mathcal{A}}$ by

$$T \mapsto \{(X, Y) \mid X \subseteq Y \text{ and } X, Y \in \sigma_{mod}(T^Y)\}$$

where

$$\begin{aligned} T^Y &= \{a_0 \leftarrow a_1, \dots, a_m \mid a_0 \leftarrow a_1, \dots, a_m, \sim a_{m+1}, \dots, \sim a_n \in T, a_{m+1}, \dots, a_n \notin Y\} \\ \sigma_{mod}(T) &= \{M \subseteq \mathcal{A} \mid \forall a_0 \leftarrow a_1, \dots, a_m, \sim a_{m+1}, \dots, \sim a_n \in T : \\ &\quad (a_1, \dots, a_m \in M \wedge a_{m+1}, \dots, a_n \notin M) \implies a_0 \in M\} \end{aligned}$$

SE-models characterize strong equivalence of stable models [?, Theorem 1]; SE-model semantics also has the intersection property [Truszczyński, 2006, Lemma 3].

Proposition 2.58. $(\mathcal{L}_{LP}, \sigma_{SE})$ is a characterization logic for $(\mathcal{L}_{LP}, \sigma_{stb})$.

Since finite-theory characterization logics are unique up to isomorphism (Theorem 2.41), there is a one-to-one-correspondence between the model sets given by Corollary 2.56 (as well as any other model set of a certain characterization logic) and sets of SE-models. More precisely, for any two logic programs $T_1, T_2 \in \mathcal{L}_{LP}$, we find $\sigma'_{stb}(T_1) \subseteq \sigma'_{stb}(T_2)$ if and only if $\sigma_{SE}(T_1) \subseteq \sigma_{SE}(T_2)$. However, please note that the set of SE-models of a finite logic program is finite, while σ'_{stb} maps logic programs to infinite model sets in general. So in the concrete case of logic programs, SE-models are much easier to work with.

2.4 Discussion

We presented a general framework for analyzing strong equivalence of knowledge representation formalisms. The framework abstracts away from all language specifics other than that knowledge bases be expressible as sets of atomic language elements. For two classes of formalisms, covered and finite-theory logics, we showed that they always possess a classical characterization logic. We called characterization logics classical because they have the intersection property (that is, the semantics of theories can always be obtained by considering the semantics of its members independently). We called characterization logics characterizing because their standard equivalence coincides with strong equivalence in the characterized formalism. As an application of our results, we obtained a first characterization logic for abstract argumentation where single AFs are interpreted as theories. This new logic complements the already existing Dung-logics which consider single AFs as building blocks and hence, theories as sets of AFs [Baumann and Brewka, 2015].

Most previous work on characterizing strong equivalence in KR that we know of focused on specific formalisms or on a handful of related formalisms, such as work on strong equivalence in logic programs under stable models [Lifschitz et al., 2001; Turner, 2001] and supported models [Truszczyński and Woltran, 2008], that also give rise to similar developments in default logic [Turner, 2001; Truszczyński, 2007] and autoepistemic logic [Truszczyński and Woltran, 2008]. By considering and exploiting formalism specifics, more fine-grained views on classical, strong and intermediate equivalence notions are possible [Eiter and Fink, 2003; Truszczyński, 2007; Woltran, 2008]. Such notions are at present not “visible” in our setting, but could be incorporated by restricting the set of theories that are allowed for expansion.

We have chosen to consider as “classical” all logics whose model function possesses the intersection property. Other characterizations might be possible when choosing consequence functions instead of model functions as starting point. For those, we considered closure operators as a special class (implying, for example, cumulativity). Other properties for future consideration come to mind – for example compactness, which is independent of the closure property and relevant to proof theory.

Our approach could be further generalized by abstracting away even from knowledge bases as sets and knowledge base expansion as set union. We could assume a language as equipped with an expansion operator \oplus under which the language is closed and then derive our results completely algebraically. This would enable us to treat, for example, abstract dialectical frameworks [Brewka and Woltran, 2010a; Brewka et al., 2013a], a quite recent non-classical KR formalism encompassing both argumentation frameworks and logic programs [Strass, 2013], for which strong equivalence has not been studied yet.

Chapter 3

Existence and Uniqueness of Argumentation Semantics

Argumentation has become one of the major fields within AI over the last two decades [Rahwan and Simari, 2009; Bench-Capon and Dunne, 2007]. In particular, Dung’s argumentation frameworks (AFs) [Dung, 1995] are widely used and act as integral concepts in several advanced argumentation formalisms. They focus entirely on conflict resolution among arguments, treating the latter as abstract items without logical structure. Hence, the only information available in AFs is the so-called attack relation that determines whether an argument is in a certain conflict with another one. As already outlined by Dung, AFs provide a formally simple basis to capture the essence of different nonmonotonic formalisms.

The subsequent sections and chapters are devoted to argumentation semantics which play the flagship role in Dung’s abstract argumentation theory. Almost all of them are motivated by an easily understandable intuition of what should be acceptable in the light of conflicts. However, although these intuitions equip us with short and comprehensible formal definitions it turned out that their intrinsic properties such as *existence and uniqueness*, *expressibility*, *replaceability* and *verifiability* are not that easily accessible. We will consider all mentioned properties w.r.t. almost all semantics available in the literature. In doing so we include two main axes: namely first, the distinction between extension-based and labelling-based versions and secondly, the distinction of different kind of argumentation frameworks such as finite or unrestricted ones.

3.1 Abstract Argumentation Theory

Phan Minh Dung introduced argumentation frameworks (AFs) as directed graphs [Dung, 1995, Definition 2]. This means, an AF $F = (A, R)$ is simply a pair consisting of a set A , usually called *set of arguments*, and a binary relation $R \subseteq A \times A$, so-called *attack relation*. In order to judge the truth of sentences like “For any AF F we have ...” or “There is no AF F , s.t. ...” we have to introduce a reference set \mathcal{U} , so-called *universe of arguments* and to require, that all possible AFs possess arguments of this set. This means, sentences as mentioned above always refer to AFs $F = (A, R)$ with $A \subseteq \mathcal{U}$. In the following we use $\mathcal{F}_{\mathcal{U}}$ as an abbreviation for the set of all AFs induced by \mathcal{U} . In order to be able to consider AFs possessing an arbitrary finite number of arguments or even infinitely many we have to request that $|\mathcal{U}| \geq |\mathbb{N}|$. No further conditions are imposed. For the rest of the thesis we assume that such a universe of arguments is given, i.e. we use the set \mathcal{U} as an arbitrary but fixed parameter. For this reason, we often use \mathcal{F} instead of $\mathcal{F}_{\mathcal{U}}$.

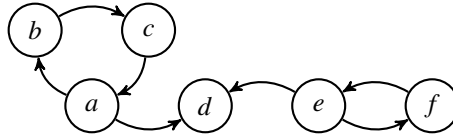
3.1.1 Important Syntactical Classes of AFs

AFs can be classified by means of syntactical properties. In the following we list some features which will be frequently used throughout the thesis. Most of them are self-explanatory and/or already well-known from graph theory [Diestel, 2006]. Most of the research in abstract argumentation typically pertains to finite AFs.

Definition 3.1. An AF $F = (A, R)$ is called

1. *empty* if $A = \emptyset$,
2. *finite* if $|A| \in \mathbb{N}$,
3. *finitary* if for any $a \in A$, $|\{b \in A \mid (b, a) \in R\}| \in \mathbb{N}$,
4. *self-loop-free* if $R \cap id_A = \emptyset$,
5. *odd-cycle-free* if there is no odd natural n , s.t. $a_0, \dots, a_n \in A$, $a_0 = a_n$ and $(a_0, a_1), \dots, (a_{n-1}, a_n) \in R$,
6. *symmetric* if $R = R^{-1}$,
7. *arbitrary* if $F \in \mathcal{F}$.

Example 3.2. The following AF is not empty, finite, finitary, self-loop-free, not odd-cycle-free, not symmetric and arbitrary.



3.1.2 Argumentation Semantics - The Flagship of Dung's Abstract Argumentation Theory

In order to formalize the notions of *existence* and *uniqueness* in the context of abstract argumentation theory we have to clarify what we precisely mean by a *semantics*. In the literature two main approaches to argumentation semantics can be found, namely so-called *extension-based* and *labelling-based* versions. The main difference is that extension-based versions return a set of sets of arguments (so-called *extensions*) for any given AF in contrast to a set of sets of n -tuples (so-called *labellings*) as in case of labelling-based approaches. However, from a mathematical point of view both kinds of semantics are instances of Definition 3.3. More precisely, extension-based versions are covered by $n = 1$ and labelling-based approaches can be obtained by setting $n \geq 2$. We use $(2^{\mathcal{U}})^n$ to denote the n -ary cartesian power of $2^{\mathcal{U}}$, i.e. $(2^{\mathcal{U}})^n = \underbrace{2^{\mathcal{U}} \times \dots \times 2^{\mathcal{U}}}_{n\text{-times}}$.

Definition 3.3. A semantics is a function $\sigma : \mathcal{F} \rightarrow 2^{(2^{\mathcal{U}})^n}$ for some natural $n \in \mathbb{N}$, s.t. $F = (A, R) \mapsto \sigma(F) \subseteq (2^A)^n$.

In this chapter we are interested in definedness statuses w.r.t. finite, finitary and arbitrary frameworks. Besides conflict-free and admissible sets (abbreviated by *cf* and *ad*) we consider a large number of mature semantics, namely naive, stage, stable, semi-stable, complete, preferred, grounded, ideal, eager semantics as well as the more exotic cf2 and stage2 semantics (abbreviated by *na*, *stg*, *stb*, *ss*, *co*, *pr*, *gr*, *il*, *eg*, *cf2* and *stg2* respectively). In the following we introduce the extension-based versions of these semantics

(indicated by \mathcal{E}_σ). Any considered semantics possesses a 3-valued labelling-based version (denoted as \mathcal{L}_σ). It is important to note that for all considered semantics we do not observe any differences between the definedness statuses of their labelling-based and extension-based versions. For the mature semantics this is due the fact that there is a one-to-one correspondence between σ -extensions and σ -labellings implying that $|\mathcal{E}_\sigma(F)| = |\mathcal{L}_\sigma(F)|$ for any AF F (for more details confer Section 5.4.1).

Before presenting the definitions we have to introduce some notational conventions. Given an AF $F = (A, R)$ and a set $E \subseteq A$. We use E_F^+ or simply, E^+ for $\{b \mid (a, b) \in R, a \in E\}$. Moreover, E_F^\oplus or simply, E^\oplus is called the *range* of E and stands for $E^+ \cup E$. We say a *attacks* b (in F) if $(a, b) \in R$. An argument a is *defended* by E (in F) if for each $b \in A$ with $(b, a) \in R$, b is attacked by some $c \in E$. Finally, $\Gamma_F : 2^A \rightarrow 2^A$ with $I \mapsto \{a \in A \mid a \text{ is defended by } I\}$ denotes the so-called *characteristic function* (of F) [Dung, 1995].

Definition 3.4. Let $F = (A, R)$ be an AF and $E \subseteq A$.

1. $E \in \mathcal{E}_{cf}(F)$ iff for no $a, b \in E$, $(a, b) \in R$,
2. $E \in \mathcal{E}_{na}(F)$ iff $E \in \mathcal{E}_{cf}(F)$ and for no $I \in \mathcal{E}_{cf}(F)$, $E \subset I$,
3. $E \in \mathcal{E}_{stg}(F)$ iff $E \in \mathcal{E}_{cf}(F)$ and there is no $I \in \mathcal{E}_{cf}(F)$, s.t. $E^\oplus \subset I^\oplus$,
4. $E \in \mathcal{E}_{stb}(F)$ iff $E \in \mathcal{E}_{cf}(F)$ and $E^\oplus = A$,
5. $E \in \mathcal{E}_{ad}(F)$ iff $E \in \mathcal{E}_{cf}(F)$ and E defends all its elements,
6. $E \in \mathcal{E}_{ss}(F)$ iff $E \in \mathcal{E}_{ad}(F)$ and there is no $I \in \mathcal{E}_{ad}(F)$, s.t. $E^\oplus \subset I^\oplus$,
7. $E \in \mathcal{E}_{co}(F)$ iff $E \in \mathcal{E}_{ad}(F)$ and for any $a \in A$ defended by E in F , $a \in E$,
8. $E \in \mathcal{E}_{pr}(F)$ iff $E \in \mathcal{E}_{ad}(F)$ and for no $I \in \mathcal{E}_{co}(F)$, $E \subset I$,
9. $E \in \mathcal{E}_{gr}(F)$ iff E is the \subseteq -least fixpoint of Γ_F ,
10. $E \in \mathcal{E}_{il}(F)$ iff $E \in \mathcal{E}_{ad}(F)$, $E \subseteq \bigcap \mathcal{E}_{pr}(F)$ and there is no $I \in \mathcal{E}_{co}(F)$ satisfying $E \subset I \subseteq \bigcap \mathcal{E}_{pr}(F)$,
11. $E \in \mathcal{E}_{eg}(F)$ iff $E \in \mathcal{E}_{ad}(F)$, $E \subseteq \bigcap \mathcal{E}_{ss}(F)$ and there is no $I \in \mathcal{E}_{co}(F)$ satisfying $E \subset I \subseteq \bigcap \mathcal{E}_{ss}(F)$.

Finally, we introduce the recursively defined cf2 and stage2 semantics [Baroni et al., 2005; Dvořák and Gaggl, 2012].

Definition 3.5. Let $F = (A, R)$ be an AF and $E \subseteq A$.

1. $E \in \mathcal{E}_{cf2}(F)$ iff
 - $E \in \mathcal{E}_{na}(F)$ if $|SCCs_F| = 1$ and
 - $\forall S \in SCCs_F(E \cap S) \in \mathcal{E}_{cf2}(F|_{UP_F(S, E)})$,
2. $E \in \mathcal{E}_{stg2}(F)$ iff
 - $E \in \mathcal{E}_{stg}(F)$ if $|SCCs_F| = 1$ and
 - $\forall S \in SCCs_F(E \cap S) \in \mathcal{E}_{stg2}(F|_{UP_F(S, E)})$.

Here $SCCs_F$ denotes the set of all strongly connected components of F , and for any $E, S \subseteq A$, $UP_F(S, E) = \{a \in S \mid \nexists b \in E \setminus S : (b, a) \in R\}$.

The following proposition summarizes well-known subset relations between the considered semantics. For two semantics σ , τ and a certain set of AFs \mathcal{C} we use $\sigma \subseteq_{\mathcal{C}} \tau$ as a shorthand for $\sigma(F) \subseteq \tau(F)$ for any AF $F \in \mathcal{C}$. The presented relations hold for both extension-based as well as labelling-based versions of the considered semantics. In the interest of readability we present the relations graphically.

Proposition 3.6. *For semantics σ and τ , $\sigma \subseteq_{\mathcal{C}} \tau$ iff there is a path of solid arrows from σ to τ in Figure 3.1. A dotted arrow indicates that the corresponding subset relation is guaranteed for finite frameworks only.*

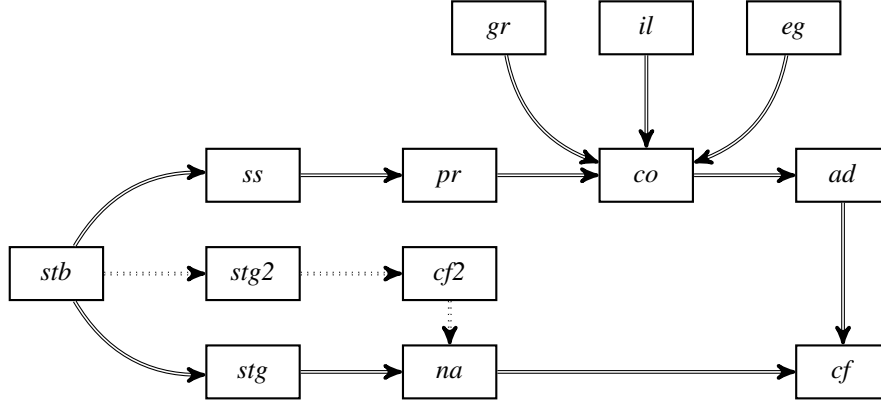
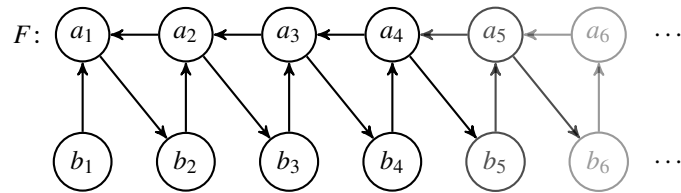


Figure 3.1: Subset Relations between Semantics

Detailed proofs can be found in [Baumann, 2014b, Proposition 2.7] as well as [Gaggl and Dvořák, 2016, Section 3.1]. Note that the shorthand $\sigma \subseteq_{\mathcal{C}} \tau$ requires that both semantics are total functions on \mathcal{C} since a framework to which one of these semantics is undefined renders the subset shorthand undefined itself. The following simple example shows that Definition 3.5 does not always provide a definite answer on whether a certain candidate set is an *cf2*-extension or *stg2*-extension, respectively. This is due to the fact that the defined recursion does not terminate necessarily in case of non-finite AFs.¹ Consequently, *stg2* and *cf2* are not total functions regarding arbitrary frameworks.

Example 3.7 (Infinite Recursion [Baumann and Spanring, 2017]). Consider the following AF $F = (A \cup B, R)$ where

- $A = \{a_i \mid i \in \mathbb{N}\}$, $B = \{b_i \mid i \in \mathbb{N}\}$ and
- $R = \{(b_i, a_i), (a_{i+1}, a_i), (a_i, b_{i+1}) \mid i \in \mathbb{N}\}$



Let $\sigma \in \{cf2, stg2\}$. We want to check whether the candidate set $E = \{b_i \mid i \in \mathbb{N}\}$ is a σ -extension. Observe that the AF F possesses two SCCs, namely one consisting of the single argument b_1 and the

¹We mention that the inventors of both semantics considered finite AFs only [Baroni et al., 2005; Dvořák and Gaggl, 2012]. In case of finite AFs any recursion will terminate no matter which candidate set is considered.

other containing the remaining arguments, i.e. $S_1 = \{b_1\}$ and $S_2 = (A \cup B) \setminus \{b_1\}$. For S_1 we end up with the base case returning a positive answer. For S_2 we have to consider the AF $F' = F|_{UP_F(S_2, E)} = F|_{(A \cup B) \setminus \{a_1, b_1\}}$ (since a_1 is attacked by $b_1 \in E \setminus S_2$) and the set $S' = E \cap S_2 = \{b_i \mid i \in \mathbb{N}, i \geq 2\}$. Obviously, determining whether S' is an σ -extension w.r.t. F' is equivalent to decide whether S is an σ -extension w.r.t. F . This means, the consideration of the candidate set E leads to infinite recursion.

3.1.3 Definedness Statuses of Argumentation Semantics

We now introduce the two main definedness statuses of argumentation semantics which capture the notions of existence and uniqueness, namely so-called *universal* and *unique definedness*. Both versions are relativized to a certain set of AFs. If clear from context, unimportant or if $\mathcal{C} = \mathcal{F}$ we will not mention explicitly the considered set of AFs.

Definition 3.8. Given a semantics σ and a set \mathcal{C} of AFs. We say that σ is

1. *universally defined* w.r.t. \mathcal{C} if $\forall F \in \mathcal{C}, |\sigma(F)| \geq 1$ and
2. *uniquely defined* w.r.t. \mathcal{C} if $\forall F \in \mathcal{C}, |\sigma(F)| = 1$.

Why are the introduced definedness statuses important? Let us consider an arbitrary logical formalism \mathcal{L} together with its semantics $\sigma_{\mathcal{L}}$. One central question is whether the semantics provides any \mathcal{L} -theory T with a formal meaning, i.e. $|\sigma_{\mathcal{L}}(T)| \geq 1$. A more demanding property than *existence* is *uniqueness*, i.e. $|\sigma_{\mathcal{L}}(T)| = 1$ for any \mathcal{L} -theory T . Clearly, these properties are interesting from several perspectives. For instance, in case of uniqueness, we observe a coincidence of sceptical and credulous reasoning modes. More precisely, if $\sigma_{\mathcal{L}}(T) = \{E\}$, then $\bigcap \sigma_{\mathcal{L}}(T) = \bigcup \sigma_{\mathcal{L}}(T) = E$. Furthermore, if a theory T is interpreted as *meaningful* if and only if $\sigma_{\mathcal{L}}(T) \neq \emptyset$, then existence might be a desired property. If the latter has to be neglected in the general case, then one further challenge is to identify sufficient properties of \mathcal{L} -theories guaranteeing their meaningfulness.

Let us come back to abstract argumentation frameworks [Dung, 1995]. Due to the practical nature of argumentation most work in the literature restricts itself to the case of finite AFs, i.e. any considered AF consists of finitely many arguments and attacks only. For this class of AFs a proof or disproof of existence or uniqueness is mostly straightforward. In the general infinite case however conducting such proofs is more intricate. It usually involves the proper use of set theoretic axioms, like the *axiom of choice* or equivalent statements. Dung already proposed the existence of preferred extensions in the case of infinite argumentation frameworks. It has later on (e.g. [Caminada and Verheij, 2010]) been pointed out that Dung has not been precise with respect to the use of principles. The existence of semi-stable extensions for finitary² argumentation frameworks was first shown by Weydert with the use of model-theoretic techniques [Weydert, 2011]. Later on, Baumann and Spanring presented a first comprehensive overview of results regarding existence and uniqueness for a whole bunch of semantics considered in the literature [Baumann and Spanring, 2015]. They provided complete or alternative proofs of already known results and contributed missing results for the infinite or finitary case. We mention two interesting results: Firstly, eager semantics is exceptional among the universally defined semantics since either there is exactly one or there are infinitely many eager extensions. Secondly, stage semantics behaves similarly to semi-stable in the sense that extensions are guaranteed as long as finitary AFs are considered. A further step forward in the systematic analysis of argumentation semantics in the infinite case was presented in [Spanring, 2015]. Spanring studied the relation between non-existence of extensions and the number of non-finitary arguments. It was shown that there are AFs where one single non-finitary argument causes a

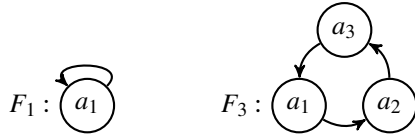
²An argument is called *finitary* if it receives finitely many attacks only. Moreover, an AF is said to be *finitary* if and only if it consists of finitary arguments only (cf. Definition 6.11).

collapse³ of semi-stable semantics. Interestingly, all known AFs which do not provide any stage extension possesses infinitely many non-finitary arguments. It is an open question whether this observation applies in general [Spanring, 2015, Conjecture 14].

3.2 Existence and Uniqueness w.r.t. Finite AFs

As a matter of fact, in order to show that a certain semantics σ is not universally defined w.r.t. a certain set \mathcal{C} it suffices to present an AF $F \in \mathcal{C}$, s.t. $\sigma(F) = \emptyset$. Contrastingly, an affirmative answer w.r.t. universal definedness requires a proof involving all AFs in \mathcal{C} . Let us consider finite AFs first. It is well-known that stable semantics does not warrant the existence of extensions/labellings even in the case of finite AFs. Witnessing examples are given by odd-cycles (cf. Example 3.9). Interestingly, in case of finite AFs we have that being odd-cycle free is sufficient for warranting at least one stable extension/labelling.⁴

Example 3.9. The following minimalistic AFs cause a collapse of stable semantics, i.e. $stb(F_1) = stb(F_3) = \emptyset$.



Observe that both frameworks do possess semi-stable, stage2 as well as stage extensions/labellings. The extensions are as follows: For any $\sigma \in \{ss, stg2, stg\}$, $\tau \in \{stg2, stg\}$, $\mathcal{E}_\sigma(F_1) = \{\emptyset\} = \mathcal{E}_{ss}(F_3)$ and $\mathcal{E}_\tau(F_3) = \{\{a_1\}, \{a_2\}, \{a_3\}\}$.

Let us consider now semi-stable semantics. Example 3.9 shows that AFs may possess semi-stable extensions even in the absence of stable extensions. Are semi-stable extensions possibly guaranteed in case of finite AFs? Consider the following explanations about the existence of semi-stable extensions taken from [Caminada, 2006]:

For every argumentation framework there exists at least one semi-stable extension. This is because there exists at least one complete extension, and a semi-stable extension is simply a complete extension in which some property (the union of itself and the arguments it defeats) is maximal.

We would like to point out two issues. Firstly, the presented explanation should not be understood as: Since any semi-stable extension is a complete one and complete semantics is universally defined we conclude that semi-stable semantics is universally defined. Accepting this kind of (false) argumentation would imply the universal definedness of stable semantics since also any stable extension is a complete one. The second issue is that the presented explanation is not precise about *why* it is guaranteed that the non-empty set of complete extensions possesses at least one range-maximal member. The following statement gives a more precise explanation [Caminada et al., 2012]:

For every (finite) argumentation framework, there exists at least one semi-stable extension. This is because there exists at least one complete extension (the grounded) and the fact that the argumentation framework is finite implies that there exist at most a finite number of complete extensions. The semi-stable extensions are then simply those complete extensions in which some property (its range) is maximal.

³The term *collapse* was firstly introduced in [Spanring, 2015] and it refers to a semantics not providing any extension/labelling for a given AF. In [Baumann and Ulbricht, 2018] the authors studied how to repair such a semantical defect.

⁴This is due to the fact that firstly, in case of finite AFs, being odd-cycle free coincides with being *limited controversial* [Dung, 1995, Definition 32] and secondly, any limited controversial AFs warrants the existence of at least one stable extensions [Dung, 1995, Corollary 36].

This means, the additional argument that we have to compare finitely many complete extensions only justifies the universal definedness of semi-stable extensions in case of finite AFs. Obviously, in case of infinite AFs we cannot expect to have finitely many complete extensions implying that this kind of argumentation is no longer valid for finitary as well as infinite AFs in general.

In the rest of this section we want to argue why all considered semantics except the stable one are universally defined in case of finite AFs.⁵ Remember that many semantics are looking for certain \subseteq -maximal elements. The main advantage in case of finiteness is that it is simply impossible to have infinite \subseteq -chains which guarantees the existence of \subseteq -maximal elements. Consider the following more detailed explanations. Given a finite AF $F = (A, R)$, i.e. $|A| = n \in \mathbb{N}$. Consequently, $1 \leq |2^A| = 2^n \in \mathbb{N}$. By definition of any extension-based semantics σ we derive $0 \leq |\mathcal{E}_\sigma(F)| \leq 2^n$ since $\mathcal{E}_\sigma(F) \subseteq 2^A$ (cf. Definition 3.3). This means, for any finite F and any semantics σ we have at least one candidate set for being a σ -extension (namely, the empty set) and at most finitely many σ -extensions. In any case, the empty set is conflict-free as well as admissible, i.e. $|\mathcal{E}_{cf}(F)|, |\mathcal{E}_{ad}(F)| \geq 1$. Furthermore, naive and preferred semantics are looking for \subseteq -maximal conflict-free or admissible sets, respectively. Since we have finitely many conflict-free as well as admissible sets only we derive the universal definedness of naive and preferred semantics in case of finite AFs. Combining $\mathcal{E}_{pr} \subseteq \mathcal{E}_{co}$ and $|\mathcal{E}_{pr}(F)| \geq 1$ yields the universal definedness of complete semantics in case of finite AFs. Moreover, since $1 \leq |\mathcal{E}_{cf}(F)|, |\mathcal{E}_{ad}(F)| \leq 2^n$ is given we obtain the universal definedness of stage and semi-stable semantics in case of finite AFs because the existence of \subseteq -range-maximal is guaranteed. Let us consider ideal and eager semantics. Candidate sets of both semantics are admissible sets being in the intersection of all preferred or semi-stable extensions, respectively. Note that there is at least one admissible set satisfying this property, namely the empty one since definitely $\emptyset \subseteq \bigcap \mathcal{E}_{pr}(F) \subseteq \mathcal{U}$ as well as $\emptyset \subseteq \bigcap \mathcal{E}_{ss}(F) \subseteq \mathcal{U}$. This means, the sets of candidates are non-empty and finite which guarantees the existence of \subseteq -maximal elements implying the universal definedness of ideal and eager semantics in case of finite AFs. The grounded extension, i.e. the \subseteq -least fixpoint of the characteristic function Γ_F , is guaranteed due to the monotonicity of Γ_F and the famous Knaster-Tarski theorem [Tarski, 1955]. Finally, even the more exotic stage2 as well as cf2 semantics are universally defined w.r.t. finite AFs. This can be seen as follows: Obviously, finitely many as well as initial SCCs are guaranteed due to finiteness. Consequently, one may start with computing stage/naive extension on these initial components and “propagate” the resulting extensions to the subsequent SCCs and so on. This procedure will definitely terminate and ends up with stage2/cf2 extensions. Apart from stable semantics we have argued that the extension-based versions of all considered semantics are universally defined w.r.t. finite AFs. In case of mature semantics, the result carry over to their labelling-based versions since any of these semantics possesses a one-to-one-correspondence between extensions and labellings. This property does not hold in case of admissible as well as conflict-free sets. However, since any admissible/conflict-free set induce at least one admissible/conflict-free labelling the result applies to their labelling versions too (cf. Section 5.4.1).

3.3 Existence and Uniqueness w.r.t. Arbitrary AFs

3.3.1 Non-well-defined Semantics

In contrast to all other semantics available in the literature, cf2 as well as stage2 semantics were originally defined recursively. The recursive schema is based on the decomposition of AFs along their strongly connected components (SCCs). Roughly speaking, the schema takes a base semantics σ and proceeds along the induced partial ordering and evaluates the SCCs according to σ while propagating relevant

⁵We mention that grounded, ideal and eager semantics are even uniquely defined w.r.t. finite AFs. This will be a by-product of Theorem 3.24, Corollary 3.23 as well as Theorem 3.26.

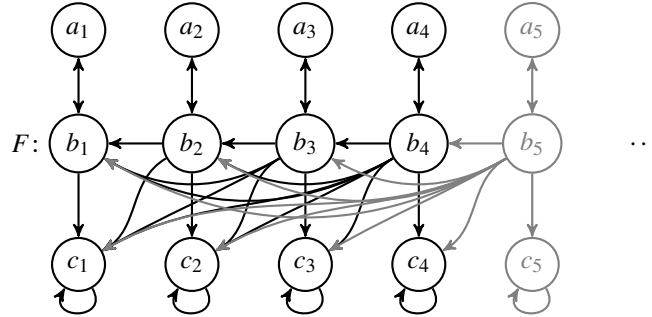
results to subsequent SCCs. This procedure defines a σ_2 semantics.⁶ Given so-called *SCC-recursiveness* [Baroni et al., 2005] we have to face some difficulties in drawing conclusions with respect to infinite AFs. Firstly, arbitrary AFs need not to possess initial SCCs which is granted for finite AFs. This makes checking whether a certain set is an σ_2 -extension more complicated and in particular, especially due to the recursive definitions not that easy to handle. Secondly, even worse, even if an AF as well as subsequent subframeworks of it possess initial SCCs there is no guarantee that any recursion will stop in finitely many steps. More precisely, as shown in Example 3.7 there might be candidate sets which lead to infinite recursion, i.e. the base case will never be considered. In [Gaggl and Dvořák, 2016, Propositions 2.12 and 3.2] the authors considered alternative non-recursive definitions of cf2 as well as stage2 semantics in case of finite AFs. It is an open question whether these definitions overcome the problem of undefinedness for arbitrary frameworks.

3.3.2 Collapsing Semantics

Dealing with finite AFs is a common as well as attractive and reasonable restriction, due to their computational nature. In the section before we have argued that apart from stable semantics all considered semantics are universally defined w.r.t. finite AFs. It is an important observation that warranting the existence of σ -extensions/labellings in case of finite AFs does not necessarily carry over to the infinite case, i.e. the semantics σ does not need to be universally defined w.r.t. arbitrary AFs. Take for instance semi-stable and stage semantics. To the best of our knowledge the first example showing that semi-stable as well as stage semantics does not guarantee extensions/labellings in case of non-finite AFs was given in [Verheij, 2003, Example 5.8.] and is picked up in the following example.

Example 3.10 (Collapse of Stage and Semi-stable Semantics). Consider the AF $F = (A \cup B \cup C, R)$ where

- $A = \{a_i \mid i \in \mathbb{N}\}, B = \{b_i \mid i \in \mathbb{N}\}, C = \{c_i \mid i \in \mathbb{N}\}$ and
- $R = \{(a_i, b_i), (b_i, a_i), (b_i, c_i), (c_i, c_i) \mid i \in \mathbb{N}\} \cup \{(b_i, b_j), (b_i, c_j) \mid i, j \in \mathbb{N}, j < i\}$

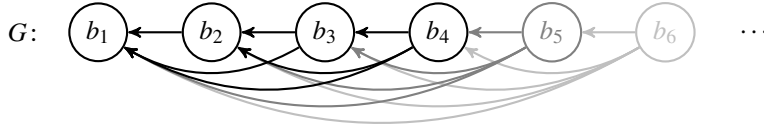


The set of preferred and naive extensions coincide, in particular $\mathcal{E}_{pr}(F) = \mathcal{E}_{na}(F) = \{A\} \cup \{E_i \mid i \in \mathbb{N}\}$ where $E_i = (A \setminus \{a_i\}) \cup \{b_i\}$. Furthermore, none of these extensions is \subseteq -range-maximal since $A^+ \subsetneq E_i^+ \subsetneq E_{i+1}^+$ for any $i \in \mathbb{N}$. In consideration of $ss \subseteq pr$ and $stg \subseteq na$ (cf. Figure 3.1) we conclude that this framework possesses neither semi-stable nor stage extensions/labellings.

In Example 3.7 we have seen that cf2 as well as stage2 semantics are not well-defined in general. This means, there are infinite AFs and candidate sets leading to an infinite recursion implying that there is no definite answer on whether such a set is an extension. However, the following example shows that even if for any candidate set a definitive decision is possible there need not to be an extension in contrast to finite AFs.

⁶Following this terminology we have to rename *cf2* semantics to *na2* semantics since its base semantics is the naive semantics and not conflict-free sets.

Example 3.11 (Collapse of Cf2 and Stage2 Semantics). Taking into account the AF $F = (A \cup B \cup C, R)$ from Example 3.10. Consider the AF $G = F|_B$, i.e. the restriction of F to B .



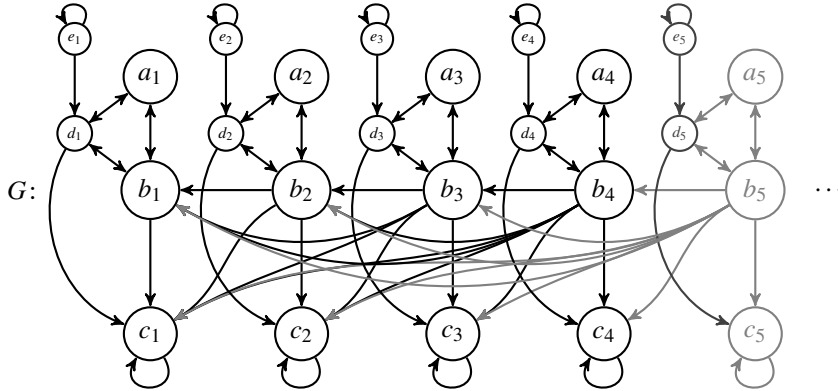
Let $\sigma \in \{cf2, stg2\}$. Obviously, any argument b_i constitutes a SCC $\{b_i\}$ which is evaluated as $\{b_i\}$ by the base semantics of σ . Consequently, \emptyset cannot be a σ -extension. Furthermore, a singleton $\{b_j\}$ cannot be a σ -extension either. The b_i 's for $i > j$ are not affected by $\{b_j\}$ and thus, the evaluation of $G|_{UP_G(\{b_i\}, \{b_j\})} = G|_{\{b_i\}} = (\{b_i\}, \emptyset)$ do not return \emptyset as required. Finally, any set containing more than two arguments would rule out at least one of them and thus, cannot be a σ -extension. Hence, $|\mathcal{E}_\sigma(G)| = |\mathcal{L}_\sigma(G)| = 0$.

In Example 3.10 we have seen an AF F without any semi-stable and stage extensions/labellings. In [Baumann and Spanring, 2015] the authors studied the question of existence-dependency between both semantics in case of infinite AFs. More precisely, they studied whether it is possible that some AF does have semi-stable but no stage extensions or vice versa, there are stage but no semi-stable extensions. The following Example 3.12 shows that stage extensions might exist even if semi-stable semantics collapses.⁷

Example 3.12 (No Semi-Stable but Stage Extensions/Labellings).

Consider again the AF F depicted in Example 3.10. Using the components of F we define $G = (A \cup B \cup C \cup D \cup E, R \cup R')$ where

- $D = \{d_i \mid i \in \mathbb{N}\}$ and $E = \{e_i \mid i \in \mathbb{N}\}$ and
- $R' = \{(a_i, d_i), (d_i, a_i), (b_i, d_i), (d_i, b_i), (d_i, c_i), (e_i, d_i), (e_i, e_i) \mid i \in \mathbb{N}\}$



In comparison to Example 3.10 we do not observe any changes as far as preferred and semi-stable semantics are concerned. In particular, $\mathcal{E}_{pr}(G) = \{A\} \cup \{E_i \mid i \in \mathbb{N}\}$ where $E_i = (A \setminus \{a_i\}) \cup \{b_i\}$ and again, none of these extensions is \subseteq -range-maximal. Hence, $\mathcal{E}_{ss}(G) = \emptyset$. Observe that we do have additional conflict-free as well as naive sets, especially the set D . Since any $e \in E$ is self-defeating and unattacked and furthermore, $D^\oplus = A \cup B \cup C \cup D$ we conclude, $\mathcal{E}_{stg}(G) = \{D\}$. Due to the one-to-one correspondence the collapse or non-collapse transfer to their labelling-based versions.

⁷The AF $G = F|_B$ depicted in Example 3.11 witnesses the reverse case. It can be checked that $\mathcal{E}_{ss}(G) = \{\emptyset\}$ and $\mathcal{E}_{stg}(G) = \emptyset$ (cf. [Baumann and Spanring, 2015, Example 2] for further explanations).

3.3.3 Universally Defined Semantics

We now turn to semantics which are universally defined w.r.t. the whole class of AFs. The first non-trivial result in this line was already proven by Dung himself, namely the universal definedness of the extension-based version of preferred semantics [Dung, 1995, Corollary 12]. He argued that the Fundamental Lemma (cf. [Dung, 1995, Lemma 10]) immediately implies that the set of all admissible sets is a complete partial order which means that any \subseteq -chain possesses a least upper bound. Then (and this was not explicitly stated in [Dung, 1995]), due to the famous *Zorn's lemma* [Zorn, 1935] the existence of \subseteq -maximal admissible sets, i.e. preferred extensions, is guaranteed.

In order to get an idea how things work in the general case we illustrate some proofs in more detail. We will see that a proof of universal definedness w.r.t. arbitrary AFs is completely different to the argumentation in case of finite ones. In order to keep this section self-contained we start with Zorn's lemma and an equivalent version of it.

Lemma 3.13 ([Zorn, 1935]). *Given a partially ordered set (P, \leq) . If any \leq -chain possesses an upper bound, then (P, \leq) has a maximal element.*

Lemma 3.14. *Given a partially ordered set (P, \leq) . If any \leq -chain possesses an upper bound, then for any $p \in P$ there exists a maximal element $m \in P$, s.t. $p \leq m$.*

Having Lemma 3.14 at hand we may easily argue that any conflict-free/admissible set is bounded by a naive/preferred extension.

Lemma 3.15. *Given $F = (A, R)$ and $E \subseteq A$,*

1. *if $E \in \mathcal{E}_{cf}(F)$, then there exists $E' \in \mathcal{E}_{na}(F)$ s.t. $E \subseteq E'$ and*
2. *if $E \in \mathcal{E}_{ad}(F)$, then there exists $E' \in \mathcal{E}_{pr}(F)$ s.t. $E \subseteq E'$.*

Proof. For $F = (A, R)$ we have the associated power set lattice $(2^A, \subseteq)$. Consider now the partially ordered fragments $\mathcal{C} = (\mathcal{E}_{cf}(F), \subseteq)$ as well as $\mathcal{A} = (\mathcal{E}_{ad}(F), \subseteq)$. In accordance with Lemma 3.14 the existence of naive and preferred supersets is guaranteed if any \subseteq -chain possesses an upper bound in \mathcal{C} or \mathcal{A} , respectively. Given a \subseteq -chain $\mathcal{E} \subseteq \mathcal{E}_{cf}(F)$ or $\mathcal{E} \subseteq \mathcal{E}_{ad}(F)$, respectively. Consider now $\bar{E} = \bigcup \mathcal{E}$. Obviously, \bar{E} is an upper bound of \mathcal{E} , i.e. for any $E \in \mathcal{E}$, $E \subseteq \bar{E}$. It remains to show that \bar{E} is conflict-free or admissible, respectively. Conflict-freeness is a finite condition. This means, if there were conflicting arguments $a, b \in \bar{E}$ there would have to be some conflict-free sets $E_a, E_b \in \bar{E}$, s.t. $a \in E_a$ and $b \in E_b$. Since \mathcal{E} is a \subseteq -chain we have $E_a \subseteq E_b$ or $E_b \subseteq E_a$ which contradicts the conflict-freeness of at least one of them. Assume now \bar{E} is not admissible. Consequently, there is some $a \in \bar{E}$ that is not defended by \bar{E} . Furthermore, there has to be an $E_a \in \mathcal{E}$, s.t. $a \in E_a$ contradicting the admissibility of $E_a \in \mathcal{E}_{ad}(F)$. \square

According to the last lemma, we may deduce the universal definedness of the extension-based versions of preferred as well as naive semantics as long as, for any AF F , the existence of at least one conflict-free or admissible set is guaranteed. This is an easy task since the empty set is conflict-free as well as admissible even in the case of arbitrary AFs. Consequently, universal definedness of both extension-based semantics is given and the same applies to their labelling-based versions due to their one-to-one correspondence.

Theorem 3.16. *Let $\sigma \in \{pr, na\}$. The semantics σ is universally defined.*

Remember that no matter which cardinality a considered AF possesses, we have that any preferred extension/labelling is a complete extension/labelling (Proposition 3.6). Thus, having the universal definedness of preferred semantics at hand we deduce that even complete semantics is universally defined w.r.t. the whole class of AFs.

Theorem 3.17. *The semantics co is universally defined.*

Let us consider now eager and ideal semantics. An eager extension is defined as the \subseteq -maximal admissible set that is a subset of each semi-stable extension. This is very similar to the definition of an ideal extension where the role of semi-stable extensions is taken over by preferred ones. On a more abstract level, both semantics are instantiations of the following schema.

Definition 3.18. Let σ be a semantics (so-called *base semantics*). We define the σ -parametrized semantics ad^σ as follows. For any AF F ,

$$\mathcal{E}_{ad^\sigma} = \max_{\subseteq} \left\{ E \in \mathcal{E}_{ad}(F) \mid E \subseteq \bigcap_{S \in \mathcal{E}_\sigma(F)} S \right\}.$$

These kind of semantics were firstly introduced in [Dvorák et al., 2011]. The authors studied general properties of these semantics in case of finite AFs with the additional restriction that the base semantics σ has to be universally defined. The following general theorem requires neither finiteness of AFs, nor any assumption on the base semantics.

Theorem 3.19. *Any σ -parametrized semantics is universally defined.*

Proof. Given an AF $F = (A, R)$ and a σ -parametrized semantics ad^σ . Consider the following set $\Sigma = \{E \in \mathcal{E}_{ad}(F) \mid E \subseteq \bigcap_{S \in \mathcal{E}_\sigma(F)} S\}$. Note that in the collapsing case, i.e. $\mathcal{E}_\sigma(F) = \emptyset$, we have: $\bigcap_{S \in \mathcal{E}_\sigma(F)} S = \{x \in \mathcal{U} \mid \forall S \in \mathcal{E}_\sigma(F) : x \in S\} = \mathcal{U}$. However, in any case $\Sigma \neq \emptyset$ since for any F , $\emptyset \in \mathcal{E}_{ad}(F)$ and obviously, $\emptyset \subseteq \bigcap_{S \in \mathcal{E}_\sigma(F)} S \subseteq \mathcal{U}$. In order to show that $\mathcal{E}_{ad^\sigma}(F) \neq \emptyset$ it suffices to prove that (Σ, \subseteq) possesses maximal elements. We will use Zorn's lemma. Given a \subseteq -chain $\mathcal{C} \in 2^\Sigma$. Consider now $\bar{E} = \bigcup \mathcal{C}$. Analogously to the proof of Lemma 3.15 we may easily show that \bar{E} is conflict-free and even admissible. Moreover, since for any $E \in \mathcal{C}$, $E \subseteq \bigcap_{S \in \mathcal{E}_\sigma(F)} S$ we deduce $\bar{E} \subseteq \bigcap_{S \in \mathcal{E}_\sigma(F)} S$ guaranteeing $\bar{E} \in \Sigma$. Now, applying Lemma 3.13, we deduce the existence of \subseteq -maximal elements in Σ , i.e. $|\mathcal{E}_{ad^\sigma}(F)| \geq 1$ concluding the proof. \square

In particular, we obtain the result for the extension-based versions of eager and ideal semantics and thus, due to the one-to-one correspondence for both labelling-based versions too.

Corollary 3.20. *Let $\sigma \in \{eg, il\}$. The semantics σ is universally defined.*

One obvious question is whether the statement above can be strengthened in the sense that both semantics are even uniquely defined w.r.t. the whole class of AFs. The following proposition, in particular the second item, shows that the unique definedness of eager semantics w.r.t. finite frameworks does not carry over to the general unrestricted case.

Proposition 3.21. *For any F we have:*

1. $ss(F) = \emptyset \Rightarrow eg(F) = pr(F)$ and
2. $ss(F) = \emptyset \Rightarrow |eg(F)| \geq \aleph_0 = |\mathbb{N}|$.

Proof. We show both assertions for the extension-based versions.

1. Given $F = (A, R)$ and let $\mathcal{E}_{ss}(F) = \emptyset$. Hence, $\bigcap_{S \in \mathcal{E}_{ss}(F)} S = \mathcal{U}$. Consequently we deduce $\mathcal{E}_{ss}(F) = \max_{\subseteq} \{E \in \mathcal{E}_{ad}(F) \mid E \subseteq \mathcal{U}\}$. This means, $\mathcal{E}_{ss}(F) = \mathcal{E}_{pr}(F)$.

2. We show the contrapositive. Assume $|\mathcal{E}_{eg}(F)| = n$ for some finite cardinal $n \in \mathbb{N}$. Due to the first statement we derive, $|\mathcal{E}_{pr}(F)| = n$. Since $ss \subseteq pr$ (cf. Proposition 3.6) we have finitely many candidates only. Furthermore, among these preferred extensions has to be at least one \subseteq -range-maximal set implying $\mathcal{E}_{ss}(F) \neq \emptyset$.

□

In a nutshell, if we observe a collapse of semi-stable semantics, then eager and preferred semantics coincide and moreover, we necessarily have infinitely many eager extensions/labellings. An AF witnessing such a behaviour can be found in Example 3.10.

3.3.4 Uniquely Defined Semantics

Although eager and ideal semantics are instances of σ -parametrized semantics we have shown the non-unique definedness (Proposition 3.21) for eager semantics only. This is no coincidence since preferred semantics, the base semantics of ideal semantics is universally defined in contrast to semi-stable semantics, the base semantics of the eager semantics. Moreover, the following theorem shows that any σ -parametrized semantics warrants the existence of exactly one extension if σ -extensions are conflict-free as well as guaranteed ([Dvorák et al., 2011, Proposition 1]).

Theorem 3.22. *Given a σ -parametrized semantics ad^σ , s.t. $\sigma \subseteq cf$ and σ is universally defined w.r.t. a class \mathcal{C} , then ad^σ is uniquely defined w.r.t. \mathcal{C} .*

Proof. Given an AF $F = (A, R)$. We already know $|\mathcal{E}_{ad^\sigma}(F)| \geq 1$ (Theorem 3.19). Hence, it suffices to show $|\mathcal{E}_{ad^\sigma}(F)| \leq 1$. Suppose, to derive a contradiction, that for some $I_1 \neq I_2$ we have $I_1, I_2 \in \mathcal{E}_{ad^\sigma}(F)$. Consequently, by Definition 3.18, $I_1, I_2 \in \mathcal{E}_{ad}(F)$ and $I_1, I_2 \subseteq \bigcap_{S \in \mathcal{E}_\sigma(F)} S$ as well as neither $I_1 \subseteq I_2$, nor $I_2 \subseteq I_1$. Obviously, $I_1 \cup I_2 \subseteq \bigcap_{S \in \mathcal{E}_\sigma(F)} S$. Since $\mathcal{E}_\sigma(F) \neq \emptyset$ and I_1 as well as I_2 has to be subsets of any σ -extension (which are conflict-free by assumption) we deduce $I_1, I_2 \in \mathcal{E}_{cf}(F)$ and thus, $I_1 \cup I_2 \in \mathcal{E}_{cf}(F)$. Furthermore, since both sets are admissible in F we derive $I_1 \cup I_2 \in \mathcal{E}_{ad}(F)$ contradicting the \subseteq -maximality of at least one of the sets I_1 and I_2 . □

Corollary 3.23. *The semantics il is uniquely defined.*

A further prominent representative of uniquely defined semantics w.r.t. the whole class of AFs is the grounded semantics. Its unique definedness was already implicitly given in [Dung, 1995]. Unfortunately, this result was not explicitly stated in the paper. Nevertheless, in [Dung, 1995, Theorem 25] it was shown that firstly, the set of all complete extensions form a complete semi-lattice w.r.t. subset relation, i.e. the existence of a \subseteq -greatest lower bound for any non-empty subset S is implied. Secondly, it was proven that the grounded extension is the \subseteq -least complete extension. Consequently, the existence of such a \subseteq -least extension is justified via setting $S = \mathcal{E}_{co}(F)$ for any given F . Alternatively, one may stick to the original definition of the grounded extension, namely as \subseteq -least fixpoint of the characteristic function Γ_F and argue that the monotonicity of Γ_F as well as the Knaster-Tarski theorem [Tarski, 1955] imply its existence.

Theorem 3.24. *The semantics gr is uniquely defined.*

3.4 Existence and Uniqueness w.r.t. Finitary AFs

Let us consider now finitary AFs, i.e. AFs where each argument receives finitely many attacks only. It was already observed by Dung itself that finitary AFs possess useful properties. More precisely, if an AF is finitary, then the characteristic function Γ is not only monotonic, but even ω -continuous [Dung, 1995,

Lemma 28] (which does not hold in case of arbitrary AFs [Baumann and Spanring, 2017, Example 1]). This implies that the least fixed point of Γ , i.e. the unique grounded extension, can be “computed” in at most ω steps by iterating Γ on the empty set (cf. [Rudin, 1976] for more details). A further advantage of finitary AFs is that for some semantics σ , the existence or even uniqueness of σ -extension is guaranteed which cannot be shown in general.

Consider again the AF F depicted in Example 3.10. In contrast to finite AFs where the existence of semi-stable as well stage extensions is guaranteed we observed a collapse of both semantics. Not that F is not finitary since, for example, the argument b_1 receives infinitely many attacks. A positive answer in case of semi-stable semantics, i.e. universal definedness w.r.t. finitary AFs was conjectured in [Caminada and Verheij, 2010, Conjecture 1] and firstly proven by Emil Weydert in [Weydert, 2011, Theorem 5.1]. Weydert proved his result in a first order logic setup using generalized argumentation frameworks. Later on, Baumann and Spanring provided an alternative proof using transfinite induction. Moreover, they showed that even stage semantics warrants the existence of at least one extension in case of finitary AFs [Baumann and Spanring, 2015, Theorem 14]. For detailed proofs we refer the reader to the mentioned scientific papers.

Theorem 3.25. *Let $\sigma \in \{ss, stg\}$. The semantics σ is universally defined w.r.t. finitary AFs.*

Applying Theorem 3.22 we derive that exactly one eager extension/labelling is guaranteed as long as the AF in question is finitary.

Theorem 3.26. *The semantics eg is uniquely defined w.r.t. finitary AFs.*

3.5 Summary of Results and Conclusion

In this section we gave an overview on the question whether certain semantics guarantee the existence or even unique determination of extensions/labellings. We have seen that these properties may vary from subclass to subclass. The following table gives a comprehensive overview over results presented in this section. The entry “ \exists ” (“ $\exists!$ ”) in row *certain* and column σ indicates that the semantics σ is universally (uniquely) defined w.r.t. the class of *certain* frameworks. No entry reflects the situation that a *certain* AF can be found which do not provide any σ -extension/labelling, i.e. σ collapses. The two question marks represent open problems. Note that we already observed that $cf2$ as well as $stage2$ semantics are not well-defined in case of finitary as well arbitrary AFs. This means, there are infinite AFs and candidate sets leading to an infinite recursion implying that there is no definite answer on whether such a set is an extension (Example 3.7). Nevertheless, even if for any candidate set a definitive decision is possible there are infinite (but non-finitary) AFs where both semantics collapse (Example 3.11). In [Baumann and Spanring, 2015, Conjecture 1] it is conjectured that this is impossible in case of finitary frameworks.

	<i>stb</i>	<i>ss</i>	<i>stg</i>	<i>cf2</i>	<i>stg2</i>	<i>pr</i>	<i>ad</i>	<i>co</i>	<i>gr</i>	<i>il</i>	<i>eg</i>	<i>na</i>	<i>cf</i>
finite		\exists	\exists	\exists	\exists	\exists	\exists	\exists	$\exists!$	$\exists!$	$\exists!$	\exists	\exists
finitary		\exists	\exists	?	?	\exists	\exists	\exists	$\exists!$	$\exists!$	$\exists!$	\exists	\exists
arbitrary						\exists	\exists	\exists	$\exists!$	$\exists!$	\exists	\exists	\exists

Table 3.1: Definedness Statues of Semantics

For a detailed complexity analysis of the associated decision problems, i.e. *Given an AF F . Is $|\sigma(F)| \geq 1$ or even, $|\sigma(F)| = 1$?* we refer the reader to [Dvořák and Dunne, 2018]. The mentioned

decisions problems are considered for finite AFs only since the input-length, i.e. the length of the formal encoding of an AF has to be finite (for finite representations of infinite AFs we refer the reader to [Baroni et al., 2013]). Due to the table above some complexity results are immediately clear. For instance, the existence problem is trivial for all considered semantics except the stable one. An upper bound for the complexity of the uniqueness problem can be obtained via the complexity of the corresponding verification problem, i.e. *Given an AF F and a set E . Is $E \in \mathcal{E}_\sigma(F)$?*. More precisely, an algorithm which decides the uniqueness problem is the following two-step procedure: first, guessing a certain set E non-deterministically and second, verifying whether this set is an σ -extension.

As already mentioned, most of the literature concentrate on finite AFs for several reasons, especially due to their computational nature. However, allowing an infinite number of arguments is essential in applications where upper bounds on the number of available arguments cannot be established a priori, such as for example in dialogues [Belardinelli et al., 2015] or modeling approaches including time or action sequences [Baumann and Strass, 2012]. Moreover, even actual infinite AFs frequently occur in the instantiation-based context. More precisely, the semantics of so-called *rule-based argumentation formalisms* (cf. Chapter 6 as well as [Prakken, 2010]) is given via the evaluation of induced Dung-style AFs. In this context, even a finite set of rules may lead to an infinite set of arguments as observed in (cf. [Caminada and Oren, 2014; Strass, 2015]).

In 2011, Baroni et al. wrote “As a matter of fact, we are not aware of any systematic literature analysis of argumentation semantics properties in the infinite case.” [Baroni et al., 2011a, Section 4.4]. Since then only few works have contributed to a better understanding of infinite AFs. In [Baroni et al., 2013] the authors studied to which extent infinite AFs can be finitely represented via formal languages and considered several decision problems within this context. In [Baumann and Spanring, 2015] a detailed study of the central properties of existence and uniqueness as presented in this chapter was given. Recently, the same authors addressed several central issues like *expressibility*, *intertranslatability* or *replaceability* (cf. Chapters 4 and 5) in the general unrestricted case [Baumann and Spanring, 2017].

Chapter 4

Expressibility of Argumentation Semantics

Given a certain logical formalism \mathcal{L} used as knowledge representation language or modelling tool in general. Depending on the application in mind, it might be interesting to know which kinds of model sets are actually expressible in \mathcal{L} ? More formally, if $\sigma_{\mathcal{L}}$ denotes the semantics of \mathcal{L} , we are interested in determining the set $\mathcal{R}_{\mathcal{L}} = \{\sigma_{\mathcal{L}}(T) \mid T \text{ is an } \mathcal{L}\text{-theory}\}$. This task, also known as *realizability* or *defineability*, highly depends on the considered formalism \mathcal{L} . Clearly, potential necessary or sufficient properties for being in $\mathcal{R}_{\mathcal{L}}$, i.e. being $\sigma_{\mathcal{L}}$ -realizable, may rule out a logic or make it perfectly appropriate for a certain application. For instance, it is well-known that in case of propositional logic any finite set of two-valued interpretations is realizable. This means, given such a finite set \mathcal{I} , we always find a set of formulae T , s.t. $\text{Mod}(T) = \mathcal{I}$. Differently, in case of normal logic programs under stable model semantics we have that any finite candidate set is realizable if it forms a \subseteq -antichain, i.e. any two sets of the candidate set have to be incomparable with respect to the subset relation. Remarkably, being such an \subseteq -antichain is not only necessary but even sufficient for realizability w.r.t. stable model semantics [Eiter et al., 2013; Strass, 2015]. One major application of realizability issues are dynamic evolvments of \mathcal{L} -theories like in case of belief revision (cf. [Alchourrón et al., 1985; Williams and Antoniou, 1998; Qi and Yang, 2008; Delgrande and Peppas, 2015; Delgrande et al., 2008; Delgrande et al., 2013; Baumann and Brewka, 2015; Diller et al., 2015] for several knowledge representation formalisms). Roughly speaking, belief revision deals with the problem of integrating new pieces of information to a current knowledge base which is represented by a certain \mathcal{L} -theory T . To this end, you are typically faced with the problem of modifying the given theory T in such a way that the revised version S satisfies $\sigma_{\mathcal{L}}(S) = M$ for some model set M . Now, before trying to do this revision in a certain minimal way it is essential to know whether M is realizable at all, i.e. $M \in \mathcal{R}_{\mathcal{L}}$.

The first formal treatment of realizability issues w.r.t. extension-based argumentation semantics was recently given by Dunne et al. [Dunne et al., 2013; Dunne et al., 2015]. They coined the term *signature* for the set of all realizable sets of extensions. The authors provided simple criteria for several mature semantics deciding whether a set of extensions is contained in the corresponding signature. For instance, two obvious necessary conditions in case of preferred semantics (as well as many other semantics) is that a candidate set S has to be non-empty, due to universal definedness of preferred semantics and second, S has to be a \subseteq -antichain, also known as *I-maximality criterion* [Baroni and Giacomin, 2007b]. However, these conditions are not sufficient implying that further requirements has to hold. In case of preferred semantics it turned out that adding the requirement of so-called *conflict-sensitivity* indeed yield a set of characterizing properties. A \subseteq -antichain S is conflict-sensitive if for each pair of distinct sets A and B from S there are at least one $a \in A$ and one $b \in B$, s.t. a and b do not occur together in any set

of S . This implies that there exists an AF F in which the set of its preferred extension coincides with $S = \{\{a, b\}, \{a, c\}, \{b, d\}, \{c, d\}\}$. Furthermore, since $\{a, b\}$ and $\{b, d\}$ are already contained in S it is impossible to realize the set $T = S \cup \{\{a, d\}\}$ under preferred semantics. From a practical point of view, such realizability insights can be used to limit the search space when enumerating preferred extensions. More precisely, applying the mentioned characterization result we obtain that not only $\{a, d\}$, but also any other set $A \subseteq \{a, b, c, d\}$ can not be a further preferred extension of a certain AF given that we already computed all sets contained in S . As a matter of fact, knowing that a certain set is realizable does not provide one automatically with a witnessing AF. Fortunately, there exist *canonical frameworks* showing realizability in a constructive fashion as shown in [Dunne et al., 2013; Dunne et al., 2015].

Later on, restricted versions of realizability were considered, namely *compact* as well as *analytic realizability* in case of extension-based semantics [Baumann et al., 2014a; Baumann et al., 2014b; Linsbichler et al., 2015; Baumann et al., 2016a]. Both versions are motivated by typical phenomena that can be observed for several semantics. First, there potentially exist arguments in a given AF that do not appear in any extension, so-called *rejected* arguments. Second, most of the argumentation semantics possess the feature of allowing *implicit conflicts*. An implicit conflict arises when two arguments are never jointly accepted although they do not attack each other. In order to understand in which way rejected arguments and implicit conflicts contribute to the expressive power of a certain semantics the notions of compact AFs as well as analytic AFs were introduced. The former kind disallows rejected arguments whereas the latter is free of implicit conflicts. It turned out that for many universally defined semantics the full range of expressiveness indeed relies on the use rejected arguments and implicit conflicts. This means, there are plenty of AFs which do not possess an equivalent AF which is in addition compact or analytic, respectively.

Recently, a first study of extension-based realizability w.r.t. arbitrary frameworks was presented in [Baumann and Spanring, 2017]. The authors compared the expressive power of several mature semantics in the unrestricted setting. Interestingly, the results reveal an intimate connection between arbitrary and finitely compact AFs in terms of expressiveness. Nevertheless, an in-depth analysis of realizability in the unrestricted setting is still missing. For instance, necessary and sufficient properties for being realizable are not considered so far.

There are only few works which have dealt with labelling-based realizability in the context of Dung-style argumentation frameworks. Dyrkolbotn showed that, as long as additional arguments are allowed any finite set of labellings is *realizable under projection* in case of preferred or semi-stable semantics [Dyrkolbotn, 2014]. In order to realize a set of labellings S under projection it suffices to come up with an AF F , s.t. its set of labellings modulo additional arguments coincide with S . The second work by Linsbichler et al. deals with the standard notion of realizability adapted to labelling-based semantics [Linsbichler et al., 2016]. The authors presented an algorithm which returns either “No” in case of non-realizability or a witnessing AF F in the positive case. Remarkably, the algorithm is not restricted to the formalism of abstract argumentation frameworks only. In fact, it can also be used to decide realizability in case of the more general abstract dialectical frameworks as well as various of its sub-classes like *bipolar* ones [Brewka and Woltran, 2010b; Brewka et al., 2013b; Baumann and Strass, 2017].

4.1 Realizability and Signatures

Let us start with the two central concepts of this section, namely *realizability* as well as *signature*. In a nutshell, we say that a certain set S is realizable under the semantics σ , if there is an AF F such that its set of σ -labellings/ σ -extensions coincides with S . Collecting all realizable sets defines the concept of a signature. In accordance with the existing literature the main part of this section is devoted to finite realizability for extension-based semantics, i.e. signatures which contain set of σ -extensions of finite AFs only. Realizability w.r.t. labelling-based semantics as well as the consideration of infinite AFs will

be briefly outlined only. Consider the following general definition of realizability in the context of abstract argumentation.

Definition 4.1. Given a semantics $\sigma : \mathcal{F} \rightarrow 2^{(2^{\mathcal{U}})^n}$ and a set $\mathcal{C} \subseteq \mathcal{F}$. A set $S \subseteq (2^{\mathcal{U}})^n$ is σ -*realizable* w.r.t. \mathcal{C} if there is an AF $F \in \mathcal{C}$, s.t. $\sigma(F) = S$.

Definition 4.2. Given a semantics σ and a set $\mathcal{C} \subseteq \mathcal{F}$. The σ -signature w.r.t. \mathcal{C} is defined as $\Sigma_{\sigma}^{\mathcal{C}} = \{\sigma(F) \mid F \in \mathcal{C}\}$.

If clear from context or unimportant we simply speak of *signatures* and write Σ without mentioning a semantics σ or set of AFs \mathcal{C} . Similarly, we say that a certain set is *realizable* instead of σ -realizable w.r.t. \mathcal{C} . Please observe that both concepts are intimately connected via the following relation: for any set S we have, S is realizable if and only if $S \in \Sigma$. Consequently, if S is not contained in Σ , then there is no framework whose extensions/labellings are exactly S . Hence, instead of searching for witnessing AFs (which might not exist) it is very attractive to find necessary as well as sufficient properties for the containment of a set S to a certain signature locally, i.e. by properties of S itself.

4.2 Signatures w.r.t. Finite AFs

We start with finite realizability. Instantiating Definitions 4.1 and 4.2 with $\mathcal{C} = \{F \in \mathcal{F} \mid F \text{ finite}\}$ formally capture the notions of realizability as well as signatures relativised to finite AFs. Consider the following definitions.

Definition 4.3. Given a semantics $\sigma : \mathcal{F} \rightarrow 2^{(2^{\mathcal{U}})^n}$. A set $S \subseteq (2^{\mathcal{U}})^n$ is *finitely σ -realizable* if there is an AF $F \in \{F \in \mathcal{F} \mid F \text{ finite}\}$, s.t. $\sigma(F) = S$.

Definition 4.4. Given a semantics σ . The finite σ -signature is defined as $\{\sigma(F) \mid F \in \mathcal{F}, F \text{ finite}\}$ abbreviated by Σ_{σ}^f .

We proceed with further notational shorthands (adjusted to the extension-based approach) which will be used throughout the whole section.

Definition 4.5 ([Dunne et al., 2015]). Given $S \subseteq 2^{\mathcal{U}}$, we use

- $Args_S$ to denote $\bigcup_{S \in \mathcal{S}} S$ and $\|S\|$ for $|Args_S|$,
- $Pairs_S$ to denote $\{(a, b) \mid \exists S \in \mathcal{S} : \{a, b\} \subseteq S\}$ and
- $dcl(S)$ to denote (the so-called *downward-closure*) $\{S' \subseteq S \mid S \in \mathcal{S}\}$

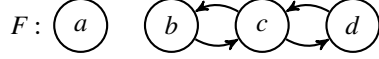
Furthermore, we say that S is an *extension-set* if $\|S\|$ is a finite cardinal.

In order to familiarize the reader with the introduced definitions we give the following example.

Example 4.6. Let $S = \{\{a\}, \{a, c\}, \{a, b, d\}\}$. Then

- $Args_S = \{a, b, c, d\}$ and $\|S\| = 4$. This means, S is an extension-set.
- $Pairs_S = \{(a, a), (b, b), (c, c), (d, d), (a, b), (a, c), (a, d), (b, d)\} \cup \{(b, a), (c, a), (d, a), (d, b)\}$
- $dcl(S) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{a, c\}, \{a, d\}, \{b, d\}, \{a, b, d\}\}$

Furthermore, since naive extensions are defined as \subseteq -maximal sets and obviously, $\{a\} \subset \{a, c\}$ we deduce that S is not *na*-realizable, i.e. $S \notin \Sigma_{\mathcal{E}_{na}}^f$. Regarding complete semantics we obtain $S \in \Sigma_{\mathcal{E}_{co}}^f$ witnessed by the following AF F .



In the following we consider the signatures of the extension-based versions of stable, semi-stable, stage, naive, preferred, complete as well as grounded semantics [Dunne et al., 2013; Dunne et al., 2015]. We provide a bunch of properties where certain subsets of them exactly matches the containment conditions for certain signatures. All properties can be decided by looking on the set in question only.

4.2.1 Semantics based on Conflict-freeness

Our starting point are semantics based on conflict-free sets. Conflict-free sets by themselves inherited their conflict-freeness to any subset of them. More formally, the downward-closure does not vary the set of conflict-free sets for a given AF. A set possessing this property is called *downward-closed*. Clearly, downward-closedness does not hold in case of admissible sets as well as any other reasonable semantics σ where conflict-freeness is just one requirement among others for being a σ -extension. Take for instance naive semantics. Naive extension are defined as \subseteq -maximal conflict-free sets. Consequently, the set of all naive extensions is a \subseteq -antichain, i.e. any two naive extensions are *incomparable* w.r.t. subset relation. This property also applies to many other semantics, such as stable and stage semantics as well as any uniquely defined semantics. However, although incomparability is a necessary condition for many considered semantics it is certainly not sufficient. Consider therefore the following example taken from [Dunne et al., 2015, Example 1].

Example 4.7. Consider the \subseteq -antichain $S = \{\{a, b\}, \{a, c\}, \{b, c\}\}$ and a semantics σ which selects its reasonable positions among the conflict-free sets, i.e. $\mathcal{E}_\sigma(F) \subseteq \mathcal{E}_{cf}(F)$ for any AF F . Now suppose there exists an AF F with $\mathcal{E}_\sigma(F) = S$. Then F must not contain attacks between a and b , a and c , and respectively b and c . This means, $\{a, b, c\} \in \mathcal{E}_{cf}(F)$. But then $\mathcal{E}_\sigma(F)$ typically contains $\{a, b, c\}$.

There are several ways to define the required property which excludes sets like S from above. It turned out that in order to characterize conflict-free based semantics like stable, stage and naive semantics a rather strong condition is required, so-called *tightness*. Roughly speaking, if an incomparable set is not tight, then there is a set $S \in \mathcal{S}$ and an argument a not belonging to S , s.t. for any $s \in S$ we find an other $S' \in \mathcal{S}$ with a and s being members of it. The idea behind the notion of being tight is simply that if an argument a does not occur in some extension S there must be a reason for that. The most simple reason one can think of is that there is a conflict between a and some $s \in S$, i.e. a and s do not occur jointly in any extension-set of \mathcal{S} or, in other words, $(a, s) \notin \text{Pairs}_{\mathcal{S}}$. In a way, this limits the multitude of incomparable elements of an extension-set.

We proceed with the formal definitions.

Definition 4.8 ([Dunne et al., 2013]). Given $S \subseteq 2^{\mathcal{U}}$. We call S

- *downward-closed* if $S = dcl(S)$,
- *incomparable* if S is a \subseteq -antichain and
- *tight* if for all $S \in \mathcal{S}$ and $a \in \text{Args}_{\mathcal{S}}$ it holds that if $S \cup \{a\} \notin \mathcal{S}$ then there exists an $s \in S$ such that $(a, s) \notin \text{Pairs}_{\mathcal{S}}$.

Please observe that for incomparable S , the premise of the tightness condition, i.e. $S \cup \{a\} \notin \mathcal{S}$, is always fulfilled. However, tightness and incomparability are independent of each other, i.e. neither tightness implies incomparability or comparability, nor incomparability implies tightness or non-tightness.

Example 4.9. Consider again the extension-set $S = \{\{a, b\}, \{a, c\}, \{b, c\}\}$ from Example 4.7. The set S is incomparable but not tight which can be seen as follows. If setting $S = \{a, b\}$ we observe $S \cup \{c\} \notin S$. Moreover, for any $s \in S$ we find an $S' \in S$, s.t. $\{s, c\} = S'$ implying that $(s, c) \in \text{Pairs}_S$. More precisely, if $s = a$, then we have $S' = \{a, c\}$ and similarly, if $s = b$ we find $S' = \{b, c\}$.

Furthermore, it can be checked that $S' = \{\{a, b\}, \{a, c\}, \{b, d\}, \{c, d\}\}$ or $S'' = S \cup \{\{a, b, c\}\}$ are witnessing examples for incomparability and tightness or tightness and comparability, respectively.

Clearly, subsets of incomparable sets are incomparable. Such a kind of inheritance does not hold in case of tight sets (cf. S and S'' as defined in Example 4.9). Nevertheless, there are non-trivial tight subsets of any tight set. For instance, in any case the set of all \subseteq -maximal elements is tight. Furthermore, if a tight set is even incomparable, then any subset of it is tight too.

In the following we present the main statements only. However, in many cases we provide some short comments indicating how to prove the statement in question. For full proofs we refer the reader to the referenced papers.

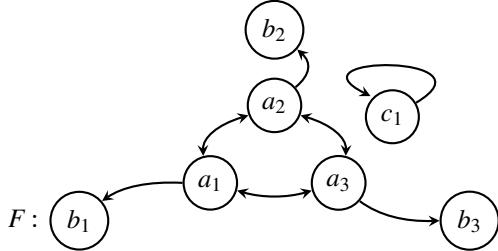
Lemma 4.10 ([Dunne et al., 2015]). *For a tight extension-set $S \subseteq 2^{\mathcal{U}}$ we have:*

1. *the \subseteq -maximal elements in S form a tight set, and*
2. *if S is incomparable then each $S' \subseteq S$ is tight.*

Note that the second statement of Lemma 4.10 implies that if the downward-closure of an incomparable extension-set S is tight, then S itself has to be tight too.

We proceed with a specific AF and check which properties apply to its different sets of extensions.

Example 4.11. Consider the following AF F .



Since c_1 is self-defeating as well as unattacked we obtain $\mathcal{E}_{stb}(F) = \emptyset$. Furthermore, $\mathcal{E}_{stg}(F) = \{\{a_1, b_2, b_3\}, \{a_2, b_1, b_3\}, \{a_3, b_1, b_2\}\}$ and $\mathcal{E}_{na}(F) = \mathcal{E}_{stg}(F) \cup \{\{b_1, b_2, b_3\}\}$. We observe,

1. $\mathcal{E}_{stb}(F), \mathcal{E}_{stg}(F)$ as well as $\mathcal{E}_{na}(F)$ are incomparable,
2. $\mathcal{E}_{stb}(F), \mathcal{E}_{stg}(F)$ as well as $\mathcal{E}_{na}(F)$ are tight and additionally,
3. $dcl(\mathcal{E}_{na}(F))$ and $dcl(\mathcal{E}_{stb}(F))$ are tight and obviously,
4. $\mathcal{E}_{stg}(F)$ and $\mathcal{E}_{na}(F)$ are non-empty.

The first and the last items are not surprising since firstly, all considered semantics satisfy the I-maximality criterion which is just another name for incomparability and secondly, in Theorems 3.25 and 3.16 we have already seen that stage extensions are guaranteed for finitary (hence, for finite) frameworks and naive semantics is even universally defined w.r.t. the whole class of AFs. This means, incomparability or non-emptiness of the mentioned sets of σ -extensions do not depend on the specific AF F , but rather apply to any finite AF. Consequently, these properties represent necessary properties regarding realizability. The tightness statements of the second and third items can be checked in a straightforward manner. We now examine that $dcl(\mathcal{E}_{stg}(F))$ is non-tight. This can be seen as follows: Firstly, $\{b_2, b_3\} \in dcl(\mathcal{E}_{stg}(F))$.

Now, for b_1 the premise of Definition 4.8 is satisfied, i.e. $\{b_1, b_2, b_3\} \notin dcl(\mathcal{E}_{stg}(F))$. Consequently, since $\{b_1, b_2\}, \{b_1, b_3\} \in dcl(\mathcal{E}_{stg}(F))$ and therefore, $(b_1, b_2), (b_1, b_3) \in Pairs_{dcl(\mathcal{E}_{stg}(F))}$ we deduce the non-tightness of $dcl(\mathcal{E}_{stg}(F))$. This means, tightness of the downward-closure of a given set can not be a necessary criterion for belonging to the stage signature.

We now present the characterization theorems for conflict-free, naive, stable as well as stage signatures. It is somehow surprising that only a few simple properties are sufficient to characterize these different signatures.

Theorem 4.12 ([Dunne et al., 2015]). *Given a set $S \subseteq 2^{\mathcal{U}}$, then*

1. $S \in \Sigma_{\mathcal{E}_{cf}}^f \iff S$ is a non-empty, downward-closed, and tight extension-set,
2. $S \in \Sigma_{\mathcal{E}_{na}}^f \iff S$ is a non-empty, incomparable extension-set and $dcl(S)$ is tight,
3. $S \in \Sigma_{\mathcal{E}_{stb}}^f \iff S$ is a incomparable and tight extension-set,
4. $S \in \Sigma_{\mathcal{E}_{stg}}^f \iff S$ is a non-empty, incomparable and tight extension-set.

We mention that a proof of the characterization theorem above requires two directions. Let us fix a certain semantics $\sigma \in \{cf, na, stb, stg\}$. The first part is to show that for any finite AF F , $\mathcal{E}_\sigma(F)$ satisfies the mentioned properties. Now, for the second part, if a certain extension-set S satisfies the properties in question, then we have to find a finite AF F , s.t. $\mathcal{E}_\sigma(F) = S$.

Let us start with the first part. It suffices to consider tightness only since downward-closedness, non-emptiness and incomparability are clear (cf. some explanations given in Example 4.11). It is easy to see that $\mathcal{E}_{cf}(F)$ is tight because if augmenting a conflict-free set S with a non-conflicting argument a yields a conflicting set, then obviously there has to be at least one element in $s \in S$, s.t. $\{a, s\}$ is conflicting. In order to prove that $dcl(\mathcal{E}_{na}(F))$ is tight, it suffice to see that $dcl(\mathcal{E}_{na}(F)) = \mathcal{E}_{cf}(F)$. Consequently, applying Lemma 4.10 we obtain the tightness of $\mathcal{E}_{na}(F)$. Furthermore, with the same lemma, we get that every $S \subseteq \mathcal{E}_{na}(F)$ is tight. In consideration of $stb \subseteq stg \subseteq na$ (Proposition 3.6) it follows that $\mathcal{E}_{stb}(F)$ as well as $\mathcal{E}_{stg}(F)$ are tight.

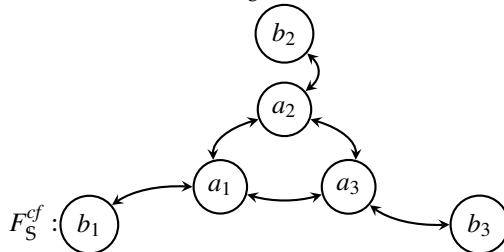
In order to show that the mentioned properties are not only necessary but even sufficient we have to come up with witnessing AFs. Consider therefore the following prototype.

Definition 4.13 ([Dunne et al., 2015]). Given an extension-set S , we define the *canonical argumentation framework* for S as

$$F_S^{cf} = (Args_S, (Args_S \times Args_S) \setminus Pairs_S).$$

The idea behind the framework is simple: we draw a relation between two arguments iff they do not occur jointly in any set $S \in \mathcal{E}_\sigma$. Consequently, for any S , F_S^{cf} is symmetric. Moreover, in any case, it is self-loop-free since $a \in Args_S$ implies $(a, a) \in Pairs_S$. Let us consider the following example.

Example 4.14. Let $S = \{\{a_1, b_2, b_3\}, \{a_2, b_1, b_3\}, \{a_3, b_1, b_2\}, \{b_1, b_2, b_3\}\}$ and consider the corresponding canonical framework F_S^{cf} .



Please note that S is non-empty, incomparable as well as possesses a tight downward-closure (cf. Example 4.11). Furthermore, F_S^{cf} realizes S under the naive semantics, i.e. $\mathcal{E}_{na}(F_S^{cf}) = S$.

The following proposition shows that this is no coincidence.

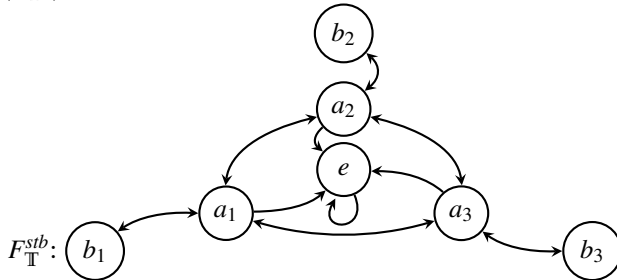
Proposition 4.15 ([Dunne et al., 2015]). *For each non-empty, incomparable extension-set S , where $dcl(S)$ is tight, $\mathcal{E}_{na}(F_S^{cf}) = S$.*

Moreover, the canonical framework can also be used as witnessing framework in case of conflict-free sets as stated in the following proposition.

Proposition 4.16 ([Dunne et al., 2015]). *For each non-empty, downward-closed and tight extension-set S , $\mathcal{E}_{cf}(F_S^{cf}) = S$.*

We proceed with stable and stage semantics. In Theorem 4.12 the only difference between the characterizations of stable and stage signatures is the non-empty requirement for stage semantics. Remember that we are dealing with finite AFs and indeed in case of this restriction stable semantics is the only semantics which does not warrant the existence of extensions (cf. Table 3.1).¹ This means, stable semantics is the only semantics which may realize the empty extension-set (which is incomparable and tight too). The final step towards concluding Theorem 4.12 is to find witnessing frameworks for any non-empty, incomparable and tight extension-sets. At first we will show that the canonical framework does not do the job in case of these semantics. More precisely, given a non-empty, incomparable as well as tight extension-set S , then the sets of stable as well as stage extensions of the canonical framework F_S^{cf} do not necessarily coincide with S .

Example 4.17. Consider again Example 4.14. We define $T = S \setminus \{b_1, b_2, b_3\}$. Please note that F_T^{cf} and F_S^{cf} are identical since $Args_S = Args_T$ and $Pairs_S = Pairs_T$. Furthermore, according to Example 4.11 we have that T is non-empty, incomparable and tight, but $\mathcal{E}_{stb}(F_T^{cf}) = \mathcal{E}_{stg}(F_T^{cf}) = \mathcal{E}_{na}(F_S^{cf}) = S \neq T$. In order to get rid of the undesired stable as well as stage extension $E = \{b_1, b_2, b_3\}$ we may simply add a new self-defeating argument e to F_S^{cf} , s.t. e is attacked by all other arguments excepting those stemming from E . The following framework F_T^{stb} illustrates this idea. Convince yourself that $\mathcal{E}_{stb}(F_T^{stb}) = \mathcal{E}_{stg}(F_T^{stb}) = T$.



The following definition generalizes the construction idea from above to arbitrary many undesired sets. The subsequent proposition states that we have indeed found witnessing examples for non-empty, incomparable and tight extension-sets as required for Theorem 4.12.

Definition 4.18 ([Dunne et al., 2015]). Given an extension-set S and its canonical framework $F_S^{cf} = (A_S^{cf}, R_S^{cf})$. Let $\mathbb{X} = \mathcal{E}_{stb}(F_S^{cf}) \setminus S$ we define

$$F_S^{stb} = (A_S^{cf} \cup \{\bar{E} \mid E \in \mathbb{X}\}, R_S^{cf} \cup \{(\bar{E}, \bar{E}), (a, \bar{E}) \mid E \in \mathbb{X}, a \in Args_S \setminus E\}).$$

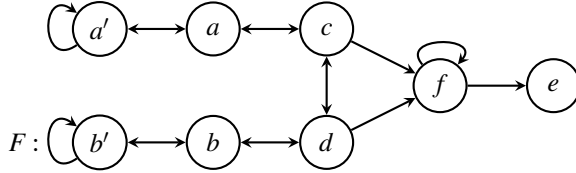
¹For instance, $F = (\{a\}, \{(a, a)\})$ yields $\mathcal{E}_{stb}(F) = \emptyset$.

Proposition 4.19 ([Dunne et al., 2015]). *For each non-empty, incomparable and tight extension-set S , $\mathcal{E}_{stb}(F_S^{stb}) = \mathcal{E}_{stg}(F_S^{stb}) = S$.*

4.2.2 Semantics based on Admissibility

Let us turn now to semantics based on admissible sets. In particular, we provide characterization theorems for the finite signatures w.r.t. admissible sets as well as preferred and semi-stable semantics. In contrast to semantics based on conflict-free sets where the notion of tightness played a decisive role (cf. Theorem 4.12) we have to introduce a new concept, so-called *conflict-sensitivity*. Conflict-sensitivity is a very basic property in the sense that it is fulfilled by almost all semantics σ (or rather, their corresponding sets of σ -extensions) available in the literature. Furthermore, it is strictly weaker than tightness, i.e. tight extension-sets are always conflict-sensitive, but not necessarily vice versa. To explain the difference between these two notions let us consider the following example taken from [Dunne et al., 2015].

Example 4.20. Consider the following framework F .



We have $\mathcal{E}_{pr}(F) = \mathcal{E}_{ss}(F) = S = \{A, B, C\} = \{\{a, b\}, \{a, d, e\}, \{b, c, e\}\}$. First, observe that S is not tight. This can be seen as follows: Obviously, $A \cup \{e\} \notin S$, but both (a, e) and (b, e) are contained in $Pairs_S$ since $\{a, e\} \subseteq B$ and $\{b, e\} \subseteq C$. This means, although $A \cup \{e\}$ is not a reasonable position w.r.t. preferred and semi-stable semantics we find witnessing extensions, namely B and C , showing that any argument in A is compatible with e , i.e. they can be accepted together. Please observe that this is not true for any two arguments in A and B or A and C , respectively. For instance, $b, d \in A \cup B$, but $(b, d) \notin Pairs_S$ as well as $a, c \in A \cup C$, but $(a, c) \notin Pairs_S$. Furthermore, the same applies to B and C , since $c, d \in B \cup C$ and $(c, d) \notin Pairs_S$.

The following definition precisely formalizes the observed property of the AF F presented in the example above.

Definition 4.21 ([Dunne et al., 2015]). A set $S \subseteq 2^{\mathcal{U}}$ is called *conflict-sensitive* if for each $A, B \in S$ such that $A \cup B \notin S$ it holds that $\exists a, b \in A \cup B : (a, b) \notin Pairs_S$.

As the name suggests, the property checks whether the absence of the union of any pair of extensions in an extension-set S is justified by a conflict indicated by S . Note that for $a, b \in A$ (likewise $a, b \in B$), $(a, b) \in Pairs_S$ holds by definition. Thus the property of conflict-sensitivity is determined by arguments $a \in A \setminus B$, $b \in B \setminus A$, for $A, B \in S$. As already indicated tightness implies conflict-sensitivity as stated in the following lemma.

Lemma 4.22 ([Dunne et al., 2015]). *Every tight extension-set is also conflict-sensitive.*

Similarly to Lemma 4.10 one may show that the set of all \subseteq -maximal elements of a conflict-sensitive set is conflict-sensitive too. Moreover, if the initial set is incomparable in addition, then even any subset of it is conflict-sensitive. Furthermore, in contrast to tight extension-sets it is possible to add the empty set to a conflict-sensitive set without losing conflict-sensitivity.²

²Note that any one-element extension-set $S \neq \{\emptyset\}$ is tight, whereas $S \cup \{\emptyset\}$ is not.

Lemma 4.23 ([Dunne et al., 2015]). *For a conflict-sensitive extension set $S \subseteq 2^{\mathcal{U}}$,*

1. *the \subseteq -maximal elements in S form a conflict-sensitive set,*
2. *if S is incomparable then each $S' \subseteq S$ is conflict-sensitive, and*
3. *$S \cup \{\emptyset\}$ is conflict-sensitive.*

Having conflict-sensitivity at hand, we are now ready to present characterization theorems for the signatures w.r.t. admissible sets as well as preferred and semi-stable semantics. Interestingly, it turns out that preferred and semi-stable semantics are equally expressive in case of finite AFs, i.e. $\Sigma_{\mathcal{E}_{pr}}^f = \Sigma_{\mathcal{E}_{ss}}^f$.

Theorem 4.24 ([Dunne et al., 2015]). *Given a set $S \subseteq 2^{\mathcal{U}}$, then*

1. $S \in \Sigma_{\mathcal{E}_{ad}}^f \iff S$ *is a conflict-sensitive extension set containing \emptyset ,*
2. $S \in \Sigma_{\mathcal{E}_{pr}}^f \iff S$ *is a non-empty, incomparable and conflict-sensitive extension set,*
3. $S \in \Sigma_{\mathcal{E}_{ss}}^f \iff S$ *is a non-empty, incomparable and conflict-sensitive extension set.*

Let us first argue that the mentioned properties are necessary conditions for being in the corresponding signature. For admissible sets it suffices to recall the following two facts: First, the empty set is admissible by definition; and second, if the union of two admissible sets is conflict-free, then the union is admissible too. In other words, if the union fails to be admissible, then there has to be a conflict proving the conflict-sensitivity of any set of admissible sets. Now, for preferred and semi-stable semantics. Non-emptiness is due to the already shown universal definedness of both semantics in case of finite AFs (cf. Table 3.1). Moreover, incomparability is clear since both semantics satisfy the I-maximality criterion (cf. [Baroni and Giacomin, 2007b]). Finally, conflict-sensitivity of sets of admissible sets transfer to sets of preferred extensions via statement 1 of Lemma 4.23 and therefore also to sets of semi-stable extensions via statement 2 of Lemma 4.23 and the fact that $ss \subseteq pr$ (Proposition 3.6).

In order to show that the mentioned properties are not only necessary but even sufficient we have to come up with witnessing AFs. In contrast to conflict-free based semantics we have to find AFs which encode the central notion of admissibility. Please note that the already introduced canonical frameworks F_S^{cf} as well as F_S^{stb} (cf. Definitions 4.13 and 4.18) do not comply with the requirements. Consider therefore the following example.

Example 4.25. Let us consider again the non-empty, incomparable as well as tight set $\mathbb{T} = \{\{a_1, b_2, b_3\}, \{a_2, b_1, b_3\}, \{a_3, b_1, b_2\}\}$ together with its corresponding canonical framework $F_{\mathbb{T}}^{stb}$ as presented in Example 4.17. Due to Lemma 4.22 we have that any tight extension-set is even conflict-sensitive and thus, \mathbb{T} satisfies the necessary requirements of Theorem 4.24. Inspecting the canonical framework reveals that $\mathcal{E}_{pr}(F_{\mathbb{T}}^{stb}) = \mathbb{T} \cup \{\{b_1, b_2, b_3\}\} \neq \mathbb{T}$. Although, $\mathcal{E}_{ss}(F_{\mathbb{T}}^{stb}) = \mathbb{T}$ one may easily check that non-empty, incomparable as well as conflict-sensitive set $S = \{\{a, b\}, \{a, d, e\}, \{b, c, e\}\}$ mentioned in Example 4.20 shows that this equality does not hold in general. Likewise, one may prove that the framework F_S^{cf} is not appropriated as a witnessing prototype for semi-stable as well as preferred semantics.

It turned out that suitable canonical AFs can be built by means of so-called *defense-formulae* as introduced in the following definition.

Definition 4.26 ([Dunne et al., 2015]). Given an extension-set S , the *defense-formula* \mathcal{D}_a^S of an argument $a \in \text{Args}_S$ in S is defined as:

$$\mathcal{D}_a^S = \bigvee_{S \in S, s.t. a \in S} \bigwedge_{s \in S \setminus \{a\}} s.$$

\mathcal{D}_a^S given as (a logically equivalent) CNF is called *CNF-defense-formula* $\mathcal{C}\mathcal{D}_a^S$ of a in S .

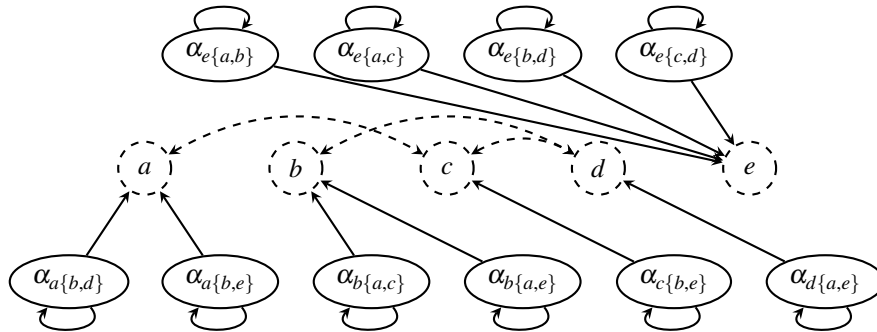
The main idea of the formula \mathcal{D}_a^S is to describe the conditions for the argument a being in an extension. Note that the variables coincide with the arguments. If S amounts to a set of admissible extensions, then each disjunct represents a set of arguments A which allows a to join in the sense that $A \cup \{a\}$ is a reasonable position w.r.t. admissible semantics. Put it differently, propositional models of $\mathcal{D}_a^S \wedge a$ represent (if considered as set of atoms) supersets of certain reasonable position. Please note that a defense-formula \mathcal{D}_a^S is tautological if and only if $\{a\} \in S$. We proceed with an example.

Example 4.27. Consider again the non-empty, incomparable as well as conflict-sensitive set $S = \{\{a, b\}, \{a, d, e\}, \{b, c, e\}\}$ stemming from Example 4.20. We obtain the following defense-formulae together with their corresponding CNF-defense-formulae (written in clause form).

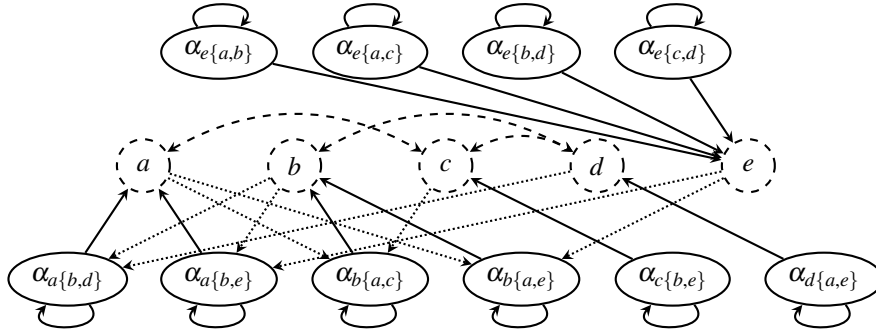
- $\mathcal{D}_a^S = b \vee (d \wedge e) \equiv (b \vee d) \wedge (b \vee e)$ and $\mathcal{CD}_a^S = \{\{b, d\}, \{b, e\}\}$
- $\mathcal{D}_b^S = a \vee (c \wedge e) \equiv (a \vee c) \wedge (a \vee e)$ and $\mathcal{CD}_b^S = \{\{a, c\}, \{a, e\}\}$
- $\mathcal{D}_c^S = b \wedge e$ and $\mathcal{CD}_c^S = \{\{b, e\}\}$
- $\mathcal{D}_d^S = a \wedge e$ and $\mathcal{CD}_d^S = \{\{a, e\}\}$
- $\mathcal{D}_e^S = (a \wedge d) \vee (b \wedge c) \equiv (a \vee b) \wedge (d \vee b) \wedge (a \vee c) \wedge (d \vee c)$ and $\mathcal{CD}_e^S = \{\{a, b\}, \{a, c\}, \{b, d\}, \{c, d\}\}$

One simple idea for the realization of a certain set S under admissible semantics is the following two-step procedure. In the first step, we construct a framework F which maintains all elements of S as conflict-free sets. This can be done via the canonical framework F_S^{cf} . In the second step, we augment the initial framework F_S^{cf} , s.t. only elements in S become admissible. The second step can be realized via adding a certain amount of additional arguments. More precisely, for any argument $a \in \text{Args}_S$ we add n self-conflicting arguments $\alpha_{aC_1}, \dots, \alpha_{aC_n}$ if $|\mathcal{CD}_a^S| = |\{C_1, \dots, C_n\}| = n$. Then, for any $i \in \{1, \dots, n\}$, α_{aC_i} attacks a and is in turn attacked by any argument in C_i . Consider therefore the following example.

Example 4.28. Again consider the extension-set $S = \{\{a, b\}, \{a, d, e\}, \{b, c, e\}\}$ and its corresponding CNF-defense-formulae as presented in Example 4.27. In accordance with the above mentioned two-step procedure we obtain the dashed AF F_S^{cf} first. Then, in view of the CNF-defense-formulae we have to add 10 additional self-defeating arguments which attacks their corresponding argument. This intermediate step is depicted below.



Let us consider the set $\{a, b\} \in S$. In order for $\{a, b\}$ to be admissible we have to add counter-attacks for the arguments $\alpha_{\{a,b,d\}}$, $\alpha_{\{a,b,e\}}$, $\alpha_{\{b,a,c\}}$ and $\alpha_{\{b,a,e\}}$. For instance, $\alpha_{\{a,b,d\}}$ is attacked by b and d and so forth. The following figure (built on top of the previous one) depicts resulting counter-attacks for the mentioned 4 arguments highlighted as densely dotted edges. For the sake of clarity we do not perform this construction for the remaining arguments.



The following definition precisely formalizes the mentioned two-step procedure.

Definition 4.29 ([Dunne et al., 2015]). Given an extension-set S , the *canonical defense-argumentation-framework* $F_S^{def} = (A_S^{def}, R_S^{def})$ extends the canonical AF $F_S^{cf} = (Args_S, R_S^{cf})$ as follows:

$$A_S^{def} = Args_S \cup \bigcup_{a \in Args_S} \{ \alpha_{a\gamma} \mid \gamma \in \mathcal{CD}_a^S \}, \text{ and}$$

$$R_S^{def} = R_S^{cf} \cup \bigcup_{a \in Args_S} \{ (b, \alpha_{a\gamma}), (\alpha_{a\gamma}, \alpha_{a\gamma}), (\alpha_{a\gamma}, a) \mid \gamma \in \mathcal{CD}_a^S, b \in \gamma \}.$$

The subsequent proposition shows that not only all elements in S become admissible in the constructed AF F_S^{def} , but rather that the set of admissible sets of F_S^{def} exactly coincides with S given that S is conflict-sensitive as well as contains the empty set.

Proposition 4.30 ([Dunne et al., 2015]). *For each conflict-sensitive extension set S where $\emptyset \in S$, it holds that $\mathcal{E}_{ad}(F_S^{def}) = S$.*

Interestingly, we may even use the canonical defense-AF to show that any non-empty, incomparable and conflict-sensitive extension-set S can be realized under the preferred semantics. This can be seen as follows: First, via Lemma 4.23 we obtain the conflict-sensitivity of $S \cup \{\emptyset\}$ since S is assumed to be conflict-sensitive. Consequently, using Proposition 4.31 we obtain $\mathcal{E}_{ad}(F_{S \cup \{\emptyset\}}^{def}) = S \cup \{\emptyset\}$. Since $F_S^{def} = F_{S \cup \{\emptyset\}}^{def}$ and due to the incomparability of S , we have $\mathcal{E}_{pr}(F_S^{def}) = S$ as stated in the following proposition.

Proposition 4.31 ([Dunne et al., 2015]). *For each non-empty, incomparable and conflict-sensitive extension-set S , it holds that $\mathcal{E}_{pr}(F_S^{def}) = S$.*

Furthermore, due to a translation result by Dvořák and Woltran we obtain that any non-empty, incomparable and conflict-sensitive extension-set S can be realized under semi-stable semantics too. More precisely, in [Dvořák and Woltran, 2011] it is shown that for any AF F exists an AF F' , s.t. $\mathcal{E}_{pr}(F) = \mathcal{E}_{ss}(F')$.

Proposition 4.32 ([Dunne et al., 2015]). *Each non-empty, incomparable and conflict-sensitive extension-set S is ss-realizable.*

4.2.3 Uniquely Defined Semantics

Let us finally turn to grounded, ideal and eager semantics. Remember that all mentioned semantics warrants the existence of exactly one extension given that the frameworks in question are finite (cf. Table 3.1). Furthermore, it is hardly surprising that this property is even sufficient for being in the corresponding signature, since any one-element extension-set $S = \{E\}$ can be realized via $F_E = (E, \emptyset)$. In particular, we obtain that all three semantics are equally expressive.

Theorem 4.33 ([Dunne et al., 2016]). *Given a set $S \subseteq 2^{\mathcal{U}}$, then*

1. $S \in \Sigma_{\mathcal{E}_{gr}}^f \iff S$ is an extension-set with $|S| = 1$,
2. $S \in \Sigma_{\mathcal{E}_{il}}^f \iff S$ is an extension-set with $|S| = 1$ and
3. $S \in \Sigma_{\mathcal{E}_{eg}}^f \iff S$ is an extension-set with $|S| = 1$.

4.2.4 Summary of Results and Further Remarks

In this subsection we provide a comprehensive overview of characterization results w.r.t. extension-based realizability in case of finite AFs. The following table collect and combine the results of the previous three subsections. The table has to be interpreted as follows: Consider a certain column σ . Then, the entries “ \times ” in rows r_1, \dots, r_n indicate that for any extension-set S , $S \in \Sigma_{\mathcal{E}_\sigma}^f \iff r_1, \dots, r_n$. Moreover, an entry “ \rightarrow ” in row r reflects the fact that the collection of the properties r_1, \dots, r_n imply property r .

	<i>cf</i>	<i>na</i>	<i>stb</i>	<i>stg</i>	<i>ad</i>	<i>pr</i>	<i>ss</i>	<i>gr</i>	<i>il</i>	<i>eg</i>
$S \neq \emptyset$	\times	\times		\times	\rightarrow	\times	\times	\rightarrow	\rightarrow	\rightarrow
$\emptyset \in S$	\rightarrow				\times					
$ S = 1$								\times	\times	\times
$dcl(S)$ is tight		\times						\rightarrow	\rightarrow	\rightarrow
S is incomparable		\times	\times	\times		\times	\times	\rightarrow	\rightarrow	\rightarrow
S is tight	\times	\rightarrow	\times	\times				\rightarrow	\rightarrow	\rightarrow
S is conflict-sensitive	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\times	\times	\times	\rightarrow	\rightarrow	\rightarrow
$dcl(S) = S$	\times									

Table 4.1: Characterizing Properties for Realizable Extension-sets

Remember that the decision whether a certain extension-set S is realizable can not be done via brute force (i.e., enumerating AFs and checking whether their extensions coincide with S) since there are no a priori bounds on the number of required arguments. Consequently, the results depicted in Table 4.1 put us in a very good position since now, the question of realizability can be decided locally, i.e. by inspecting the set in question itself. Moreover, all mentioned properties can be checked in polynomial time

w.r.t. the size of the extensions as shown in [Dunne et al., 2015, Theorem 6]. For the majority of the properties tractability is immediately apparent. The only exception is tightness of the downward-closure of a given extension-set S since its size is not polynomially bounded in the size of S (cf. [Dunne et al., 2015, Proposition 12] for a way out of this problem).

By inspecting the respective properties as depicted in Table 4.1, we can immediately put the signatures of different semantics in relation to each other. The following theorem includes the signature w.r.t. complete semantics in addition. The reason why we did not include complete semantics in our considerations is simply that a precise characterization of the complete signature is still an open problem. Nevertheless, certain necessary properties are already found [Dunne et al., 2015, Proposition 4] justifying items 3 and 4 of the following theorem.

Theorem 4.34 ([Dunne et al., 2015]). *The following relations hold*

1. $\Sigma_{\mathcal{E}_{na}}^f \subset \Sigma_{\mathcal{E}_{stg}}^f \subset \Sigma_{\mathcal{E}_{ss}}^f = \Sigma_{\mathcal{E}_{pr}}^f$,
2. $\Sigma_{\mathcal{E}_{stb}}^f = \Sigma_{\mathcal{E}_{stg}}^f \cup \{\emptyset\}$,
3. $\Sigma_{\mathcal{E}_{cf}}^f \subset \Sigma_{\mathcal{E}_{ad}}^f \subset \Sigma_{\mathcal{E}_{co}}^f$,
4. $\Sigma_{\mathcal{E}_{\sigma}}^f \subset \Sigma_{\mathcal{E}_{\tau}}^f$ where $\sigma \in \{gr, il, eg\}$, $\tau \in \{na, stb, stg, pr, ss, co\}$ and
5. $\{S \cup \{\emptyset\} \mid S \in \Sigma_{\mathcal{E}_{pr}}^f\} \subset \Sigma_{\mathcal{E}_{ad}}^f$.

The following Venn-diagram provides a compact overview of subset relations between the considered signatures. A bordered area represents a set of extension-sets. The outer ellipse $\mathcal{E}_{\mathcal{S}} = \{S \subseteq 2^{\mathcal{U}} \mid S \text{ is an extension set}\}$ stands for the set of all extension-sets over \mathcal{U} . Clearly, all other signatures are subsets of $\mathcal{E}_{\mathcal{S}}$ by definition. Furthermore, we use $\{\{\emptyset\}\}$ or $\{\emptyset\}$ the set consisting of the single extension-set (realizable by all considered semantics) or the set containing the empty extension-set (realizable by stable semantics only), respectively. The right side of Figure 4.1 shows signatures of semantics providing only incomparable extension-sets. The intersection of these signatures with $\Sigma_{\mathcal{E}_{co}}^f$ exactly coincides with $\Sigma_{\mathcal{E}_{gr}}^f$ as well as $\Sigma_{\mathcal{E}_{il}}^f$ and $\Sigma_{\mathcal{E}_{eg}}^f$ which contain all extension-sets S with $|S| = 1$. Moreover, the only extension-set they have in common with the signatures of conflict-free and admissible sets is the extension-set containing the empty extension. This fact causes the “missing” intersection in the middle of Figure 4.1.

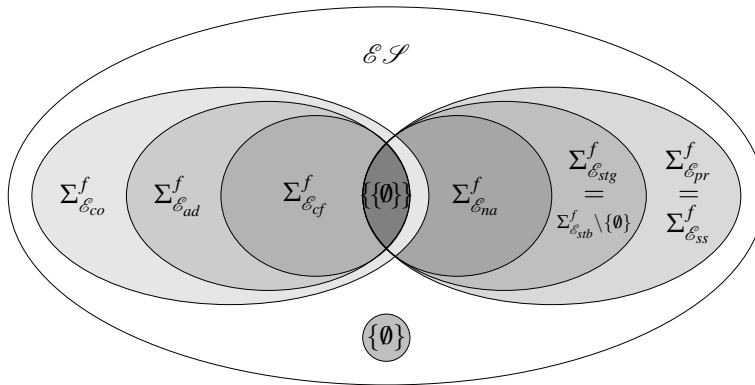


Figure 4.1: Subset Relations between Finite Signatures

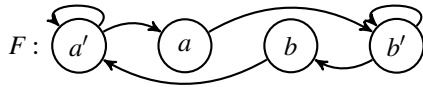
Finally, we want to mention that all considered finite signatures, apart from the complete signature, are closed under non-empty intersections. More precisely, if two finitely σ -realizable sets \mathbb{S} and \mathbb{T} possess a non-empty intersection, then $\mathbb{S} \cap \mathbb{T}$ is finitely σ -realizable too. This feature is mainly due to the fact that subsets of incomparable and tight as well as incomparable and conflict-sensitive sets maintain these properties (cf. Lemmas 4.10 and 4.23).

Theorem 4.35 ([Dunne et al., 2015]). *Let $\sigma \in \{cf, ad, na, stb, stg, pr, ss\}$. For any two finite AFs F_1, F_2 exists an finite AF F , s.t. $\mathcal{E}_\sigma(F) = \mathcal{E}_\sigma(F_1) \cap \mathcal{E}_\sigma(F_2)$ given that $\mathcal{E}_\sigma(F_1) \cap \mathcal{E}_\sigma(F_2) \neq \emptyset$.³*

4.3 Signatures w.r.t. Finite, Compact AFs

So far we considered realizability without any restriction (apart from finiteness) for witnessing AFs. This means, realizing AFs may contain *rejected* arguments, i.e. arguments which do not appear in any extension. Rejected arguments are natural ingredients in typical argumentation scenarios and it is a priori completely unclear in which ways rejected arguments contribute to the expressibility of a particular semantics. In order to have a handle for analyzing the effect of rejected arguments, the class of *compact* AFs and its induced signatures were introduced and studied [Baumann et al., 2014a; Baumann et al., 2014b; Baumann et al., 2016a]. An AF is compact with respect to a semantics σ , if it does not contain rejected arguments, i.e. each of its arguments appears in at least one σ -extension. Now, the main question is whether it is possible to get rid of rejected arguments without changing the outcome? or, in other words: Under which circumstances can AFs be transformed into equivalent compact ones? Note that studying compactness is far from being an academic exercise since there is a fundamental computational significance: When searching for extensions, arguments span the search space, since extensions are to be found among the subsets of the set of all arguments. Hence the more arguments, the larger the search space. Compact AFs are argument-minimal since none of the arguments can be removed without changing the outcome, thus leading to a minimal search space.

Let us first have a brief look on the naive semantics, which is defined as \subseteq -maximal conflict-free sets: Here, it is rather easy to see that any AF can be transformed into an equivalent compact AF by just removing all self-defeating arguments. In other words, the same outcome (in terms of the naive extensions) can be achieved by a simplified AF without rejected arguments. This means, naive semantics does not lose expressive power if we stick to compact AFs. However, it is not hard to find semantics where this coincidence does not hold implying that for such semantics the full range of expressiveness indeed relies on the concepts of rejected arguments. Consider therefore the following non-compact AF F .



Let us consider admissible sets. We obtain $\mathbb{S} = \mathcal{E}_{ad}(F) = \{\emptyset, \{a, b\}\}$. Obviously, any attempt of realizing \mathbb{S} with a compact AF $G = (\{a, b\}, R)$ is doomed to failure since if $\{a, b\}$ is admissible in G we necessarily obtain the admissibility of $\{a\}$ as well as $\{b\}$ proving $\mathbb{S} \neq \mathcal{E}_{ad}(G)$. It was one main result in [Baumann et al., 2014a] to show that the finite, compact signatures w.r.t. stable, preferred, semi-stable, and stage semantics are strict subsets of their corresponding finite signatures. This means, in case of those semantics, sticking to finite, compact AFs implies a loss of expressive power.

³The prerequisite of a non-empty intersection can be dropped in case of stable semantics.

4.3.1 Central Definitions and Preliminary Observations

In the following we formally introduce the central notions of *compact argumentation frameworks*, *compact realizability* as well as *compact signatures*. As already stated, the main idea behind compact AFs is the absence of rejected arguments. For labelling-based semantics σ (i.e., a semantics returning n -tuples) we assume that the first component of their associated σ -labellings are interpreted as *acceptable sets of arguments* in analogy to σ -extensions in case of extension-based semantics. This means, if a certain argument occur in no first component of given σ -labellings we classify it as *rejected*. For a given labelling \mathcal{L} we use \mathcal{L}^1 to refer to its first component.

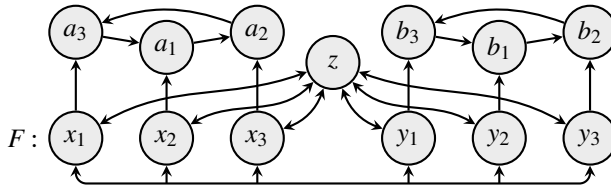
Definition 4.36. Given a semantics $\sigma : \mathcal{F} \rightarrow 2^{(2^{\mathcal{U}})^n}$. An AF $F = (A, R)$ is compact for σ (or simply, σ -compact) if $\text{Args}_{\mathcal{E}_\sigma(F)} = A$ (in case of $n = 1$) or $\text{Args}_{\{\mathcal{L}^1 \mid \mathcal{L} \in \mathcal{L}_\sigma(F)\}} = A$ (for $n \geq 2$), respectively.

Although extension-based and labelling-based semantics are formally different semantics (according to Definition 3.3) we often speak of the extension-based version or labelling-based version of a certain semantics. This can be formally justified for the considered semantics since there is a close relationship between both versions (cf. Facts 5.38 and 5.39 for some formal relations). The following fact shows that for all considered semantics σ there is no need to distinguish between σ -compactness w.r.t. the extension-based version of σ and σ -compactness w.r.t. the labelling-based version of σ . As an aside, such a coincidence does not require a one-to-one correspondence between the extension-based and labelling-based version of a semantics σ . It suffices that any σ -extension induces a σ -labelling and vice versa in such a way that accepted arguments are preserved (cf. statements 1 and 2 of Fact 5.38).

Fact 4.37. For any $\sigma \in \{stb, ss, stg, cf2, stg2, pr, ad, co, gr, il, eg, na, cf\}$ and any⁴ AF F we have: F is compact for \mathcal{E}_σ iff F is compact for \mathcal{L}_σ .

In the following we use CAF_σ for AFs compact for σ . Moreover, we use CAF_σ^f to indicate that the considered frameworks are finite in addition. It is intuitively clear that there are AFs F being σ -compact without being τ -compact for two different semantics σ and τ . The following example firstly presented in [Baumann et al., 2014a, Figure 1] provides us with a witnessing framework.

Example 4.38. Consider the following AF F .⁵



The preferred extensions of F are $\mathcal{E}_{pr}(F) = \{\{z\}, \{x_1, a_1\}, \{x_2, a_2\}, \{x_3, a_3\}, \{y_1, b_1\}, \{y_2, b_2\}, \{y_3, b_3\}\}$, meaning that F is *pr*-compact ($F \in CAF_{pr}^f$) since each argument occurs in at least one preferred extension. On the other hand observe that $\mathcal{E}_{ss}(F) = \mathcal{E}_{pr}(F) \setminus \{\{z\}\}$ and $\mathcal{E}_{stg}(F) = \{\{x_i, a_i, b_j\}, \{y_i, b_i, a_j\} \mid 1 \leq i, j \leq 3\}$, i.e. z is not contained in any semi-stable or stage extension. Therefore F is neither compact for semi-stable nor compact for stage semantics (i.e. $F \notin CAF_{ss}^f$ and $F \notin CAF_{stg}^f$).

How are the different sets of compact AFs related? We start with an easy observation.

Lemma 4.39 ([Baumann et al., 2016a]). For any two semantics σ and τ such that for each AF F , for every $S \in \mathcal{E}_\sigma(F)$ there is some $S' \in \mathcal{E}_\tau(F)$ with $S \subseteq S'$, we have $CAF_\sigma \subseteq CAF_\tau$.

⁴Indeed, no finiteness restriction is required here.

⁵The construct in the lower part of the figure represents symmetric attacks between each pair of distinct arguments. We will make use of this style in illustrations throughout the whole section.

Note that $\sigma \subseteq \tau$ is a special case of the premise of Lemma 4.39. Thus, $CAF_\sigma \subseteq CAF_\tau$, whenever $\sigma \subseteq \tau$ (see Figure 3.1 for an overview). Strict subset relations have to be proven by providing a witnessing AF as presented in Example 4.38. Moreover, $CAF_{pr} = CAF_{co} = CAF_{ad}$ as well as $CAF_{na} = CAF_{cf}$ is justified by Lemma 3.15 and the fact that $pr \subseteq co \subseteq ad$ and $na \subseteq cf$. Finally, in case of the uniquely defined grounded and ideal semantics we have, $F = (A, R)$ is compact if and only if $R = \emptyset$. This in turn implies that F is compact for stable semantics. This means, $CAF_{gr} = CAF_{il} \subset CAF_{stb}$. Remember that eager semantics is uniquely defined w.r.t. finitary AFs only (Theorem 3.26, Example 3.10). Consequently, we may conclude $CAF_{gr}^f = CAF_{eg}^f$ only. Although, the majority of the results do not require the finiteness restriction we present the following theorem in terms of finite AFs. Detailed proofs for the relations between stable, semi-stable, preferred, stage and naive semantics can be found in [Baumann et al., 2016a, Theorem 2].

Theorem 4.40. *The following relations hold:*

1. $CAF_{gr}^f = CAF_{il}^f = CAF_{eg}^f$,
2. $CAF_{pr}^f = CAF_{co}^f = CAF_{ad}^f$,
3. $CAF_{na}^f = CAF_{cf}^f$,
4. $CAF_{gr}^f \subset CAF_{stb}^f \subset CAF_{ss}^f \subset CAF_{pr}^f \subset CAF_{na}^f$,
5. $CAF_{stb}^f \subset CAF_{sig}^f \subset CAF_{na}^f$ and
6. $CAF_{sig}^f \not\subseteq CAF_\sigma^f$ as well as $CAF_\sigma^f \not\subseteq CAF_{sig}^f$ for any $\sigma \in \{pr, ss\}$.

The following figure concisely summarizes all relations mentioned in the theorem above. Directed arrows between two boxes have to be interpreted as strict subset relations between the mentioned sets of compact AFs in these boxes.

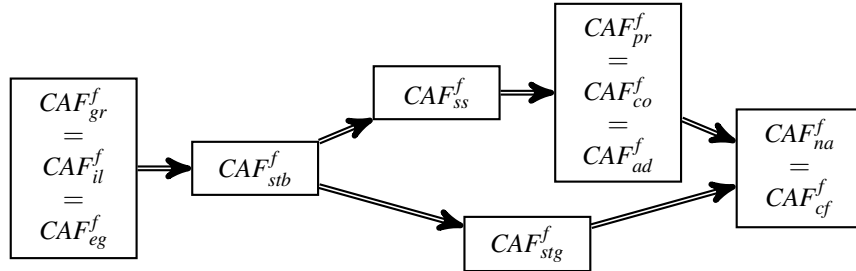


Figure 4.2: Subset Relations between Finite, Compact AFs

Instantiating Definitions 4.1 and 4.2 with $\mathcal{C} = CAF_\sigma^f$ formalize the notions of realizability as well as signatures relativised to finite, compact AFs. Consider the following definitions.

Definition 4.41. Given a semantics $\sigma : \mathcal{F} \rightarrow 2^{(2^{\mathcal{U}})^n}$. A set $S \subseteq (2^{\mathcal{U}})^n$ is *finitely, compactly σ -realizable* if there is an AF $F \in CAF_\sigma^f$, s.t. $\sigma(F) = S$.

Definition 4.42. Given a semantics σ . The finite, compact σ -signature is defined as $\{\sigma(F) \mid F \in CAF_\sigma^f\}$ abbreviated by $\Sigma_\sigma^{f,c}$.

It is clear that $\Sigma_\sigma^{c,f} \subseteq \Sigma_\sigma^f$ holds for any semantics σ , i.e. finite, compact realizability implies finite realizability. In the following we shed light on the question whether the mentioned subset relation is strict for a given semantics? In other words, we answer the question whether we indeed lose expressive power if sticking to compact AFs.

4.3.2 The Loss or Stability of Expressive Power

Let us consider the uniquely defined grounded, ideal and eager semantics first. We already stated that a set S is realizable w.r.t. these semantics if and only if S is an one-element extension-set if considering finite AFs (Theorem 4.33). Furthermore, it is immediate that an extension-set $S = \{E\}$ can be compactly realized via $F_E = (E, \emptyset)$. This means, these semantics do not lose expressive power if we restrict ourselves to compact AFs. Furthermore, the attentive reader may have noticed that the canonical argumentation framework F_S^{cf} , which was used as a witnessing framework for conflict-free sets and naive semantics (cf. Definition 4.13 as well as Propositions 4.15 and 4.16), does not involve further artificial arguments. Thus, it verifies finite, compact realizability and shows that there is no expressive loss in case of conflict-free sets and naive semantics. For the other considered semantics, namely admissible, stable, stage, semi-stable, preferred as well as complete semantics we have to accept a strict weaker expressibility if we stick to compact AFs. In order to prove that in case of these semantics the full range of expressiveness indeed relies on the concept of rejected arguments we have to come up with witnessing extension-sets. Consider therefore the following example.

Example 4.43. The extension-set $S = \{\{a, b\}, \{a, d, e\}, \{b, c, e\}\}$ is realizable under preferred as well as semi-stable semantics (cf. Example 4.20 for a realizing non-compact framework). Let $\sigma \in \{pr, ss\}$. Now suppose there exists an AF $F = (\{a, b, c, d, e\}, R)$, s.t. $\mathcal{E}_\sigma(F) = S$. Since $\{a, d, e\}, \{b, c, e\} \in S$ and $\sigma \subseteq cf$ we conclude that there is no attack in R involving e , i.e. e is an isolated argument in F . But then, e is contained in each σ -extension of F contradicting $\{a, b\} \in S$. In Summary, $S \in \Sigma_{\mathcal{E}_\sigma}^f \setminus \Sigma_{\mathcal{E}_\sigma}^{f,c}$.

For further witnessing extension-sets we refer the reader to [Baumann et al., 2016a, Propositions 35 and 57] and proceed with the main theorem.

Theorem 4.44. *It holds that*

1. $\Sigma_{\mathcal{E}_\sigma}^{f,c} = \Sigma_{\mathcal{E}_\sigma}^f$ for $\sigma \in \{cf, na, gr, il, eg\}$, and
2. $\Sigma_{\mathcal{E}_\sigma}^{f,c} \subset \Sigma_{\mathcal{E}_\sigma}^f$ for $\sigma \in \{ad, stb, stg, ss, pr, co\}$.

In both cases we may benefit of characterization theorems for finite signatures (cf. Theorems 4.12, 4.24 and 4.33). If both signatures are identical (first item), then necessary and sufficient properties for being finitely σ -realizable immediately carry over to finite, compact σ -realizability. If we observe a strict subset relation (second item), then we obtain at least necessary properties for being in the finite, compact σ -signature.

Theorem 4.45. *Given a set $S \subseteq 2^{\mathcal{U}}$, then*

1. $S \in \Sigma_{\mathcal{E}_{cf}}^{f,c} \iff S$ is a non-empty, downward-closed and tight extension set,
2. $S \in \Sigma_{\mathcal{E}_{na}}^{f,c} \iff S$ is a non-empty, incomparable extension set and $dcl(S)$ is tight,
3. $S \in \Sigma_{\mathcal{E}_{gr}}^{f,c} \iff S$ is an extension set with $|S| = 1$,
4. $S \in \Sigma_{\mathcal{E}_{il}}^{f,c} \iff S$ is an extension set with $|S| = 1$,
5. $S \in \Sigma_{\mathcal{E}_{eg}}^{f,c} \iff S$ is an extension set with $|S| = 1$ and
6. $S \in \Sigma_{\mathcal{E}_{stb}}^{f,c} \implies S$ is an incomparable and tight extension set,
7. $S \in \Sigma_{\mathcal{E}_{stg}}^{f,c} \implies S$ is a non-empty, incomparable and tight extension set,

8. $S \in \Sigma_{\mathcal{E}_{ad}}^{f,c} \implies S$ is a conflict-sensitive extension set containing \emptyset ,
9. $S \in \Sigma_{\mathcal{E}_{pr}}^{f,c} \implies S$ is a non-empty, incomparable and conflict-sensitive extension set,
10. $S \in \Sigma_{\mathcal{E}_{ss}}^{f,c} \implies S$ is a non-empty, incomparable and conflict-sensitive extension set.

4.3.3 Comparing Finite, Compact Signatures and Final Remarks

In the following we relate the finite, compact signatures of the semantics under consideration to each other. Recall that for finite signatures it holds that $\Sigma_{\mathcal{E}_{na}}^f \subset \Sigma_{\mathcal{E}_{stg}}^f = (\Sigma_{\mathcal{E}_{stb}}^f \setminus \{\emptyset\}) \subset \Sigma_{\mathcal{E}_{ss}}^f = \Sigma_{\mathcal{E}_{pr}}^f$ (cf. Figure 4.1). This picture changes dramatically when considering the relationships between finite, compact signatures as depicted in Figure 4.3 (incomparable semantics only) and formally stated in Theorem 4.46. The dashed areas represent particular intersections for which the question of existence of extension-sets is still an open question.

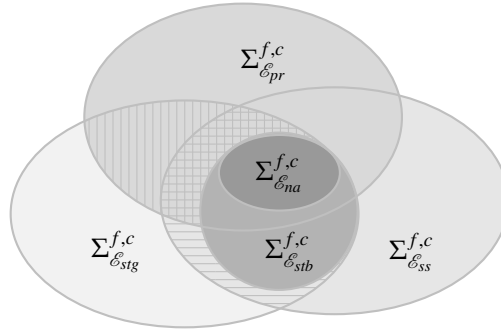


Figure 4.3: Subset Relations between Finite, Compact Signatures

We proceed with an enumeration of relationships between finite, compact signature including further semantics like conflict-free and admissible sets as well as grounded, ideal, eager and complete semantics. For formal proofs we refer the interested reader to [Baumann et al., 2016a, Theorem 36, Proposition 58].

Theorem 4.46. *The following relations hold:*

1. $\Sigma_{\mathcal{E}_{\sigma}}^{f,c} \subset \Sigma_{\mathcal{E}_{na}}^{f,c} \subset \Sigma_{\mathcal{E}_{\tau}}^{f,c}$ for $\sigma \in \{gr, il, eg\}$ and $\tau \in \{stb, stg, ss, pr\}$,
2. $\Sigma_{\mathcal{E}_{stb}}^{f,c} \subset \Sigma_{\mathcal{E}_{\sigma}}^{f,c}$ for $\sigma \in \{stg, ss\}$,
3. $\Sigma_{\mathcal{E}_{cf}}^{f,c} \subset \Sigma_{\mathcal{E}_{ad}}^{f,c}$,
4. $\Sigma_{\mathcal{E}_{co}}^{f,c} \setminus \Sigma_{\mathcal{E}_{\sigma}}^{f,c} \neq \emptyset$ and $\Sigma_{\mathcal{E}_{\sigma}}^{f,c} \setminus \Sigma_{\mathcal{E}_{co}}^{f,c} \neq \emptyset$ for $\sigma \in \{cf, ad\}$,
5. $\Sigma_{\mathcal{E}_{pr}}^{f,c} \setminus (\Sigma_{\mathcal{E}_{stb}}^{f,c} \cup \Sigma_{\mathcal{E}_{ss}}^{f,c} \cup \Sigma_{\mathcal{E}_{stg}}^{f,c}) \neq \emptyset$,
6. $\Sigma_{\mathcal{E}_{stg}}^{f,c} \setminus (\Sigma_{\mathcal{E}_{stb}}^{f,c} \cup \Sigma_{\mathcal{E}_{pr}}^{f,c} \cup \Sigma_{\mathcal{E}_{ss}}^{f,c}) \neq \emptyset$,
7. $\Sigma_{\mathcal{E}_{stb}}^{f,c} \setminus \Sigma_{\mathcal{E}_{pr}}^{f,c} \neq \emptyset$,
8. $(\Sigma_{\mathcal{E}_{pr}}^{f,c} \cap \Sigma_{\mathcal{E}_{ss}}^{f,c}) \setminus (\Sigma_{\mathcal{E}_{stb}}^{f,c} \cup \Sigma_{\mathcal{E}_{stg}}^{f,c}) \neq \emptyset$ and

$$9. \Sigma_{\mathcal{E}_{ss}}^{f,c} \setminus (\Sigma_{\mathcal{E}_{stb}}^{f,c} \cup \Sigma_{\mathcal{E}_{pr}}^{f,c} \cup \Sigma_{\mathcal{E}_{stg}}^{f,c}) \neq \emptyset.$$

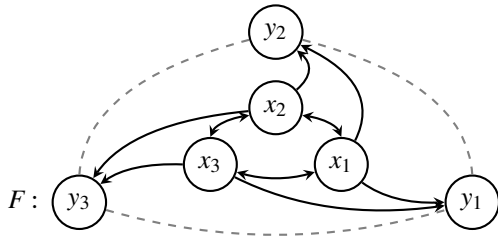
Comparing the results on expressiveness of the considered semantics as stated in Theorems 4.34 and 4.46 we observe notable differences. When allowing rejected arguments, preferred and semi-stable semantics are equally expressive and at the same time strictly more expressive than stable and stage semantics. As we have seen, this property does not carry over to the compact setting (with the exceptions $\Sigma_{\mathcal{E}_{stb}}^{f,c} \subset \Sigma_{\mathcal{E}_{ss}}^{f,c}$ and $\Sigma_{\mathcal{E}_{stb}}^{f,c} \subset \Sigma_{\mathcal{E}_{stg}}^{f,c}$) where signatures become incomparable.

Finally, regarding the open issues represented as dashed areas in Figure 4.3. More precisely, it is an open problem whether there are extension-sets lying in the intersection between $\Sigma_{\mathcal{E}_{pr}}^{f,c}$ (resp. $\Sigma_{\mathcal{E}_{ss}}^{f,c}$) and $\Sigma_{\mathcal{E}_{stg}}^{f,c}$ but outside of $\Sigma_{\mathcal{E}_{stb}}^{f,c}$. In [Baumann et al., 2016a] it is conjectured that such extension-sets do not exist.

Conjecture 4.47 ([Baumann et al., 2016a]). *It holds that $\Sigma_{\mathcal{E}_{pr}}^{f,c} \cap \Sigma_{\mathcal{E}_{stg}}^{f,c} \subset \Sigma_{\mathcal{E}_{stb}}^{f,c}$ and $\Sigma_{\mathcal{E}_{ss}}^{f,c} \cap \Sigma_{\mathcal{E}_{stg}}^{f,c} = \Sigma_{\mathcal{E}_{stb}}^{f,c}$.*

4.4 Signatures w.r.t. Finite, Analytic AFs

We now turn to a further phenomenon, so-called *implicit conflicts* which can be frequently observed in typical argumentation scenarios. Consider therefore the following AF F .



Let us consider stable semantics. Please note that any x_i is jointly acceptable with one specific y_j . More precisely, $\mathcal{E}_{stb}(F) = \{\{x_1, y_3\}, \{x_2, y_1\}, \{x_3, y_2\}\}$ implying that we do not have any rejected arguments, i.e. F is stable compact. What can be said about the two pairs of arguments x_1 and x_2 as well as y_1 and y_2 ? First of all, both pairs represent a semantical conflict in F since neither of those pairs occur together in any stable extension. In case of x_1 and x_2 , the conflict is even a syntactical one since both arguments attack each other in contrast to the pair consisting of y_1 and y_2 . This difference leads to the distinction between syntactically underlined *explicit conflicts* and syntactically unfounded *implicit* ones (indicated by dashed lines). In order to understand how implicit conflicts contribute to the expressiveness of a certain semantics, the set of *analytic* AFs and its induced signatures were introduced and studied [Linsbichler et al., 2015; Baumann et al., 2016a]. An analytic framework, i.e. a framework which is free of implicit conflicts maximizes the information on conflicts. One main question is: under which circumstances an arbitrary framework can be transformed into an equivalent analytic one? This question is interesting from a theoretical as well as practical point of view. On the one hand, analytic frameworks are natural candidates for normal forms of AFs, and on the other maximizing the number of explicit conflicts might help argumentation systems to evaluate AFs more efficiently.

Let us consider again the extension-set $S = \{\{x_1, y_3\}, \{x_2, y_1\}, \{x_3, y_2\}\}$ stemming from the AF F depicted above. Replacing the dashed arrows with symmetric attacks in F shows that S can be analytically realized under stable semantics. Interestingly, this is no coincidence, since it was shown that in case of stable semantics any AF can be transformed into an equivalent analytical one. However, in general it is not that easy to make implicit conflicts explicit since there are frameworks where any suitable transformation requires the use of additional arguments as shown in [Linsbichler et al., 2015].

4.4.1 Central Definitions and Preliminary Observations

In this section we consider the central notions of *analytic argumentation frameworks*, *analytic realizability* as well as *analytic signatures*. In order to define analytic AF we have to differentiate between the concept of an attack (as a syntactical element) and the concept of a conflict (with respect to the evaluation under a given semantics). More precisely, if two arguments cannot be accepted together, i.e. no reasonable position contain them jointly as elements, we say that these arguments are in conflict. If this conflict is syntactically underlined by an attack between them, we call this conflict explicit, otherwise implicit. Now, an analytic framework is an AF which simply does not contain any implicit conflicts. Consider the following definition.

Definition 4.48. Given a semantics $\sigma : \mathcal{F} \rightarrow 2^{(2^{\mathcal{U}})^n}$, an AF $F = (A, R)$ and two arguments $a, b \in A$. We say that

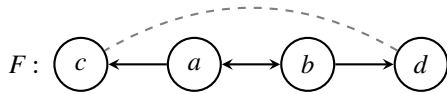
1. a and b are in *conflict* for σ if $(a, b) \notin \text{Pairs}_{\mathcal{E}_\sigma(F)}$ (in case of $n = 1$) or $(a, b) \notin \text{Pairs}_{\{\mathcal{L}^i \mid \mathcal{L} \in \mathcal{L}_\sigma(F)\}}$ (for $n \geq 2$), respectively,
2. the conflict is *explicit* w.r.t. σ if $(a, b) \in R$ or $(b, a) \in R$, otherwise *implicit*,
3. the AF F is *analytic* for σ (or σ -analytic) if all conflicts are explicit.

Please notice that Definition 4.48 does not require a and b to be different arguments. In particular, an argument that is not contained in any reasonable position is in conflict with itself. This conflict is explicit if the argument is self-attacking and implicit otherwise. Furthermore, for all considered semantics σ we observe there is no need to distinguish between σ -analyticity w.r.t. the extension-based version of σ and σ -analyticity w.r.t. the labelling-based version of σ (similarly as in case of σ -compactness as stated in Fact 4.37). Please note that this coincidence is justified for any semantics σ whenever $\mathcal{E}_\sigma(F) = \{\mathcal{L}^i \mid \mathcal{L} \in \mathcal{L}_\sigma(F)\}$ is guaranteed.

Fact 4.49. For any $\sigma \in \{stb, ss, stg, cf2, stg2, pr, ad, co, gr, il, eg, na, cf\}$ and any AF F we have: F is analytic for \mathcal{E}_σ iff F is analytic for \mathcal{L}_σ .

In the following we denote the set of all σ -analytic AFs as XAF_σ . To indicate that the frameworks under consideration are finite we use XAF_σ^f . We proceed with an example illustrating the new definitions.

Example 4.50. As a simple example consider the following AF F depicted below.



For $\sigma \in \{stb, pr, ss, stg\}$ we have $\mathcal{E}_\sigma(F) = \{\{a, d\}, \{b, c\}\}$. Observe that there is only one implicit conflict, namely the conflict between the arguments c and d , denoted by a dashed line. Hence, F is not σ -analytic, i.e. $F \notin XAF_\sigma^f$. However, since $\mathcal{E}_{na}(F) = \mathcal{E}_\sigma(F) \cup \{\{c, d\}\}$ we have that F is *na*-analytic, i.e. $F \in XAF_{na}^f$.

As indicated in Example 4.50 the sets of analytic AFs can differ for different semantics. Just like in case of compact AFs (cf. Lemma 4.39) one may easily verify the following lemma which allows to obtain a plenty of subset relations between sets of analytic AFs.

Lemma 4.51 ([Baumann et al., 2016a]). For any two semantics σ and τ such that for each AF F , for every $S \in \mathcal{E}_\sigma(F)$ there is some $S' \in \mathcal{E}_\tau(F)$ with $S \subseteq S'$, we have $XAF_\sigma \subseteq XAF_\tau$.

In line with the existing literature we restrict our considerations to finite AFs. Regarding universal (but not uniquely) defined semantics we obtain the same relations as in case of compact AFs (see explanations below Lemma 4.39). In any case we have $XAF_{gr}^f \subseteq XAF_{il}^f \subseteq XAF_{eg}^f$ since ideal semantics accepts more arguments than grounded semantics and eager semantics is even more credulous than ideal semantics. Furthermore, $XAF_{eg}^f \subseteq XAF_{ss}^f$ because the unique eager extension is contained in all semi-stable extension by definition and moreover, semi-stable semantics guarantees reasonable positions in case of finite AFs. Now, let $F = (A, R)$ be analytic w.r.t. eager semantics and $\mathcal{E}_{eg}(F) = \{E\}$. We deduce that all arguments in $A \setminus E$ have to be self-defeating. Consequently, its corresponding (conflict-free) base semantics (cf. Definition 3.18) warrants exactly one extension for F . More precisely, $\mathcal{E}_{ss}(F) = \{E\}$. Finally, due to the self-conflicting arguments and the admissibility of E we obtain $\mathcal{E}_{pr}(F) = \{E\}$ and thus, $\mathcal{E}_{il}(F) = \{E\}$ showing that F is even analytic w.r.t. ideal semantics, i.e. $XAF_{il}^f = XAF_{eg}^f$. The AF $F = (\{a, b\}, \{(a, b), (b, a), (b, b)\})$ proves that a similar result in case of grounded and ideal semantics does not hold. Detailed proofs for the relations between stable, semi-stable, preferred, stage and naive semantics can be found in [Baumann et al., 2016a, Theorem 4].

Theorem 4.52. *The following relations hold:*

1. $XAF_{gr}^f \subset XAF_{il}^f = XAF_{eg}^f \subset XAF_{ss}^f$,
2. $XAF_{pr}^f = XAF_{co}^f = XAF_{ad}^f$,
3. $XAF_{na}^f = XAF_{cf}^f$,
4. $XAF_{stb}^f \subset XAF_{ss}^f \subset XAF_{pr}^f \subset XAF_{na}^f$,
5. $XAF_{stb}^f \subset XAF_{stg}^f \subset XAF_{na}^f$,
6. $XAF_{stg}^f \not\subseteq XAF_{\sigma}^f$ and $XAF_{\sigma}^f \not\subseteq XAF_{stg}^f$ for any $\sigma \in \{pr, ss\}$,
7. $XAF_{\sigma}^f \not\subseteq XAF_{\tau}^f$ and $XAF_{\tau}^f \not\subseteq XAF_{\sigma}^f$ for any $\sigma \in \{gr, il, eg\}$, $\tau \in \{stb, stg\}$.

The following figure summarizes all relation in a compact way. Similarly to Figure 4.2, a directed arrow between two boxes has to be interpreted as strict subset relation between the mentioned sets of analytic AFs therein.

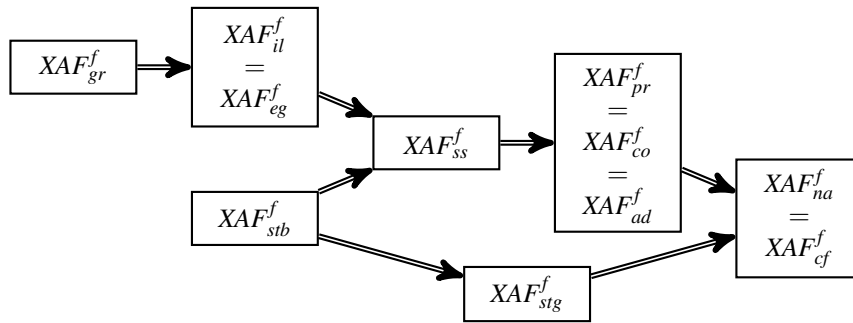


Figure 4.4: Subset Relations between Finite, Analytic AFs

At this point we want to mention that although Figures 4.2 and 4.4 look very similar we have that compactness and analyticity are sufficiently distinct properties. More precisely, apart from the uniquely

defined semantics as well as naive semantics and conflict-free sets no subset relations between the sets of compact and analytic frameworks can be stated in general. Sticking to self-loop-free AFs allows one to draw further relations such as analyticity implies compactness for any considered semantics. The main reason for this general relation is that rejected arguments has to be self-defeating in case of analytic frameworks. A selection of proofs of relations listed below can be found in [Baumann et al., 2016a, Proposition 5-8].

Proposition 4.53. *Given an AF F , then*

1. $CAF_\sigma^f \subset XAF_\sigma^f$ for $\sigma \in \{gr, il, eg, na, cf\}$,
2. $CAF_\sigma^f \not\subseteq XAF_\sigma^f$ and $XAF_\sigma^f \not\subseteq CAF_\sigma^f$ for $\sigma \in \{ad, stb, ss, pr, stg, co\}$.

If F is self-loop-free in addition, then

3. $F \in XAF_\sigma^f$ and $F \in CAF_\sigma^f$ for $\sigma \in \{na, cf\}$,
4. $F \in XAF_\sigma^f \iff F \in CAF_\sigma^f$ for $\sigma \in \{gr, il, eg\}$ and
5. $F \in XAF_\sigma^f \implies F \in CAF_\sigma^f$ for $\sigma \in \{ad, stb, ss, pr, stg, co\}$.

We now precisely formalize the notions of realizability as well as signatures relativised to finite, analytic AFs. This can be formally done via instantiating Definitions 4.1 and 4.2 with $\mathcal{C} = XAF_\sigma^f$.

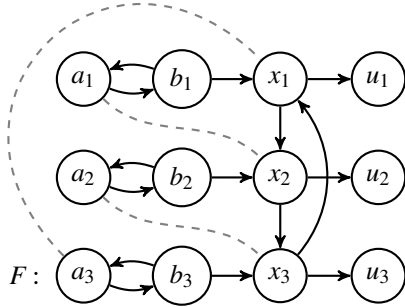
Definition 4.54. Given a semantics $\sigma : \mathcal{F} \rightarrow 2^{(2^{\mathcal{U}})^n}$. A set $S \subseteq (2^{\mathcal{U}})^n$ is *finitely, analytically σ -realizable* if there is an AF $F \in XAF_\sigma^f$, s.t. $\sigma(F) = S$.

Definition 4.55. Given a semantics σ . The finite, analytic σ -signature is defined as $\{\sigma(F) \mid F \in XAF_\sigma^f\}$ abbreviated by $\Sigma_\sigma^{f,x}$.

4.4.2 The Loss or Stability of Expressive Power

Clearly, every set in the finite, analytic signature of a semantics is also contained in the finite signature. Remember that in case of compact AFs we do not lose any expressive power if considering the uniquely defined grounded, ideal and eager semantics as well as naive semantics and conflict-free sets (Theorem 4.44). These equal expressiveness results carry over to analytic AFs and moreover, even stable and stage semantics may realize the same sets. For instance, consider again the non-analytic AF F as introduced in Example 4.50. One may easily verify that adding an attack from c to d or vice versa yields an AF F' analytic for stable semantics which does not change the set of stable extensions. However, in general it is not that easy to make implicit conflicts explicit but it was shown that the use of additional arguments indeed allows one to turn any finite framework in an analytical one without changing the set of stable or stage extensions, respectively [Baumann et al., 2016a, Proposition 28, Theorem 29]. For the sake of completeness, we mention that it was an open question for a while, known as *Explicit Conflict Conjecture* [Baumann et al., 2014a], whether it is possible, under stable semantics, to translate a given AF into an equivalent analytic one without adding further arguments. In [Baumann et al., 2016a] the conjecture was refuted for stable and even stage semantics. For the remaining semantics, i.e. admissible, semi-stable, preferred and complete semantics the conjecture does not hold either since in case of these semantics we even have that the finite, analytic signature is a strict subset of the corresponding finite one. This means, the full range of expressiveness indeed relies on the use of implicit conflicts. Consider the following example firstly presented in [Baumann et al., 2016a, Example 6].

Example 4.56. Take into account the AF $F = (A, R)$ as depicted below.



Formally, we have

$$A = \{a_i, b_i, x_i, u_i \mid i \in \{1, 2, 3\}\} \text{ and}$$

$$R = \{(a_i, b_i), (b_i, a_i), (b_i, x_i), (x_i, u_i) \mid i \in \{1, 2, 3\}\} \cup \{(x_1, x_2), (x_2, x_3), (x_3, x_1)\}.$$

Regarding the extension-based version of preferred semantics we obtain the set $S = \mathcal{E}_{pr}(F) = \{S_a, S_b, A_1, A_2, A_3, B_1, B_2, B_3\}$ with

$$S_a = \{a_1, a_2, a_3\}$$

$$S_b = \{b_1, b_2, b_3, u_1, u_2, u_3\}$$

$$A_1 = \{a_2, a_3, b_1, x_2, u_1, u_3\}$$

$$B_1 = \{a_1, b_2, b_3, x_1, u_2, u_3\}$$

$$A_2 = \{a_1, a_3, b_2, x_3, u_1, u_2\}$$

$$B_2 = \{a_2, b_1, b_3, x_2, u_1, u_3\}$$

$$A_3 = \{a_1, a_2, b_3, x_1, u_2, u_3\}$$

$$B_3 = \{a_3, b_1, b_2, x_3, u_1, u_2\}$$

We observe three implicit conflicts indicated by dashed lines. Consequently, F is not analytic w.r.t. preferred semantics. Moreover, we claim that S is not analytically *pr*-realizable at all. For a contradiction we assume that there exists an AF $G \in XAF_{pr}^f$, s.t. $\mathcal{E}_{pr}(G) = S$. We now investigate this hypothetical AF G . The main idea is to show that if the conflict between a_1 and x_2 is made explicit, then $S \neq \mathcal{E}_{pr}(G)$. First, note that G contains at least all arguments in A since $Args_S = A$. Due to A_3 and B_3 we deduce that $S_a \cup \{u_2\}$ is conflict-free in G . Furthermore, due to A_1 , the admissibility of S_a in G and the assumption that all conflicts has to be explicit, we infer that a_1 attacks x_2 . Moreover, in consideration of S , it is easy to see that x_2 is the only possible attacker of u_2 among $Args_S$. This implies that S_a defends u_2 against all arguments in $Args_S$. Finally, any additional argument $z \notin Args_S$ in G must be attacked by S_a since G is analytic w.r.t. preferred semantics and S_a must be admissible. This causes $S_a \cup \{u_2\}$ to be admissible in G and hence, S_a cannot be preferred in G . In summary, any AF realizing S has to be non-analytic for preferred semantics, i.e. $S \in \Sigma_{\mathcal{E}_{pr}}^f \setminus \Sigma_{\mathcal{E}_{pr}}^{f,x}$.

We proceed with the main theorem comparing finite signatures with their corresponding analytical ones.

Theorem 4.57 ([Baumann et al., 2016a]). *It holds that*

1. $\Sigma_{\mathcal{E}_{\sigma}}^{f,x} = \Sigma_{\mathcal{E}_{\sigma}}^f$ for $\sigma \in \{cf, na, gr, il, eg, stb, stg\}$, and
2. $\Sigma_{\mathcal{E}_{\sigma}}^{f,x} \subset \Sigma_{\mathcal{E}_{\sigma}}^f$ for $\sigma \in \{ad, ss, pr, co\}$.

In the following we present characterization theorems for finite, analytic signatures or at least necessary properties for being finitely, analytically realizable. All results can be verified via combining the main theorem above as well as the already presented characterization theorems for finite signatures, namely Theorems 4.12, 4.24 and 4.33.

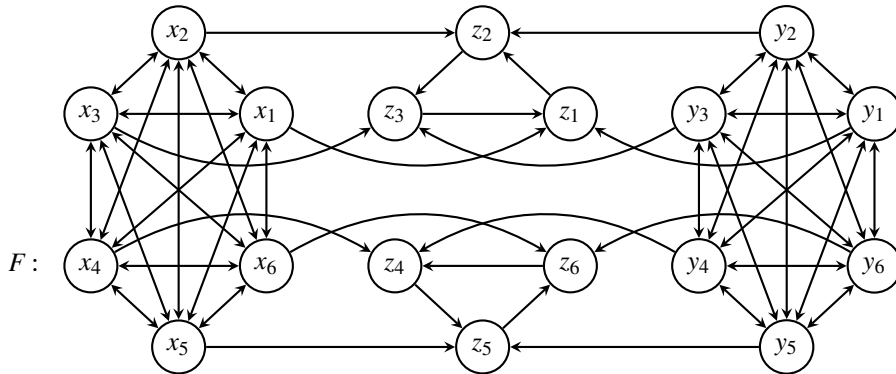
Theorem 4.58. *Given a set $S \subseteq 2^{\mathcal{U}}$, then*

1. $S \in \Sigma_{\mathcal{E}_{cf}}^{f,x} \iff S$ is a non-empty, downward-closed and tight extension set,
2. $S \in \Sigma_{\mathcal{E}_{na}}^{f,x} \iff S$ is a non-empty, incomparable extension set and $dcl(S)$ is tight,
3. $S \in \Sigma_{\mathcal{E}_{gr}}^{f,x} \iff S$ is an extension set with $|S| = 1$,
4. $S \in \Sigma_{\mathcal{E}_{il}}^{f,x} \iff S$ is an extension set with $|S| = 1$,
5. $S \in \Sigma_{\mathcal{E}_{eg}}^{f,x} \iff S$ is an extension set with $|S| = 1$,
6. $S \in \Sigma_{\mathcal{E}_{stb}}^{f,x} \iff S$ is a incomparable and tight extension set,
7. $S \in \Sigma_{\mathcal{E}_{stg}}^{f,x} \iff S$ is a non-empty, incomparable and tight extension set and
8. $S \in \Sigma_{\mathcal{E}_{ad}}^{f,x} \implies S$ is a conflict-sensitive extension set containing \emptyset ,
9. $S \in \Sigma_{\mathcal{E}_{pr}}^{f,x} \implies S$ is a non-empty, incomparable and conflict-sensitive extension set,
10. $S \in \Sigma_{\mathcal{E}_{ss}}^{f,x} \implies S$ is a non-empty, incomparable and conflict-sensitive extension set.

4.4.3 Comparing Finite, Analytic Signatures and Final Remarks

So far we have compared finite signatures and finite, analytic signatures for the semantics under consideration. We have seen, for example, that preferred and semi-stable semantics can realize strictly more when allowing the use of implicit conflicts, while this is not the case for stable and stage semantics. In the following we relate the finite, analytic signatures of all considered semantics. Remember that we observed a considerable variety in the relations between incomparable semantics if sticking from finite to finite, compact signatures (cf. Figures 4.1 and 4.3). However, in the analytic case we have slight differences only as illustrated in Figure 4.7 (for incomparable semantics) and formally stated in Theorem 4.60. For instance, preferred and semi-stable signatures do not coincide anymore as shown by the following example taken from [Baumann et al., 2016a, Figure 9, Proof of Theorem 34].

Example 4.59. Consider the following AF F as depicted below.



The preferred extension of F can be compactly presented via a cyclic successor functions. More precisely, if $s(1) = 2, s(2) = 3, s(3) = 1$ and $s(4) = 5, s(5) = 6, s(6) = 4$, then $\mathcal{E}_{pr}(F) = S = S_0 \cup S_1 \cup S_2$ with

$$\begin{aligned} S_0 &= \left\{ \{x_i, y_j, z_{s(i)}, z_{s(j)}\} \mid i \in \{1, 2, 3\}, j \in \{4, 5, 6\} \text{ or } i \in \{4, 5, 6\}, j \in \{1, 2, 3\} \right\}, \\ S_1 &= \left\{ \{x_i, y_i, z_{s(i)}\} \mid i \in \{1, 2, 3, 4, 5, 6\} \right\} \text{ and} \\ S_2 &= \left\{ \{x_i, y_{s(i)}, z_{s(s(i))}\}, \{x_{s(i)}, y_i, z_{s(s(i))}\} \mid i \in \{1, 2, 3, 4, 5, 6\} \right\}. \end{aligned}$$

This means, F is *pr*-analytic and therefore, $S \in \Sigma_{\mathcal{E}_{pr}}^{f,x}$. We show now that $S \notin \Sigma_{\mathcal{E}_{ss}}^{f,x}$. Assume that there is some $G = (B, S) \in XAF_{ss}^f$ with $\mathcal{E}_{pr}(G) = S$. We take a look at S_1 and more specifically $\{x_1, y_1, z_2\} \in S_1$. Now we need an explicit conflict between x_1 and x_4 , but in the selected set only x_1 can possibly defend against this attack, hence $(x_1, x_4) \in S$. The same argument works for x_1 and x_3 as well as z_2 and z_3 , meaning that also $(x_1, x_3), (z_2, z_3) \in S$. For symmetry reasons $\{(x_i, x_j), (x_j, x_i), (y_i, y_j), (y_j, y_i) \mid i \in \{1, 2, 3\}, j \in \{4, 5, 6\}\} \subseteq S$ and $\{(x_{s(i)}, x_i), (z_i, z_{s(i)}) \mid i \in \{1, 2 \dots 6\}\} \subseteq S$.

We take a look at S_2 and more specifically $\{x_1, y_2, z_3\} \in S_2$. As there should be an explicit conflict between x_1 and x_2 with only x_1 possibly defending this extension against x_2 we need $(x_1, x_2) \in S$. Further as in this set only y_2 and z_3 can possibly attack z_2 we have the set $\{y_2, z_3\}$ attacking z_2 . For symmetry reasons $\{(x_i, x_{s(i)}), (y_i, y_{s(i)}) \mid i \in \{1, 2 \dots 6\}\} \subseteq S$ and each set $\{x_i, z_{s(i)}\}, \{y_i, z_{s(i)}\}$ for $i \in \{1, 2 \dots 6\}$ attacks z_i .

Finally we take a look at S_0 and specifically the set $I = \{x_1, y_4, z_2, z_5\} \in S_0$. Since I necessarily is an admissible extension in an analytic AF we have that I attacks all rejected arguments. By the above observations we now have that I even attacks all arguments not being member of I in G , which means that I is a stable extension and stable semantics and semi-stable semantics thus coincide on G . But then, with $J = \{x_1, y_1, z_2\} \in S_1$ not being in conflict with for instance z_4 we have that J can not be a stable or semi-stable extension in G concluding $S \notin \Sigma_{\mathcal{E}_{ss}}^{f,x}$.

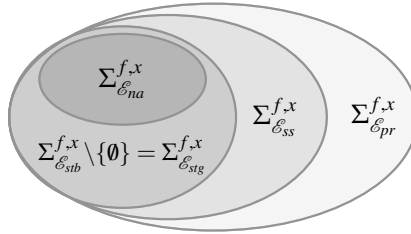


Figure 4.5: Subset Relations between Finite, Analytic Signatures

Theorem 4.60 ([Dunne et al., 2015]). *The following relations hold:*

1. $\Sigma_{\mathcal{E}_\sigma}^{f,x} \subset \Sigma_{\mathcal{E}_{na}}^{f,x} \subset \Sigma_{\mathcal{E}_{stg}}^{f,x} \subset \Sigma_{\mathcal{E}_{ss}}^{f,x} \subset \Sigma_{\mathcal{E}_{pr}}^{f,x}$ for $\sigma \in \{gr, il, eg\}$,
2. $\Sigma_{\mathcal{E}_{stb}}^{f,x} = \Sigma_{\mathcal{E}_{stg}}^{f,x} \cup \{\emptyset\}$,
3. $\Sigma_{\mathcal{E}_{cf}}^{f,x} \subset \Sigma_{\mathcal{E}_{ad}}^{f,x}$ and
4. $\Sigma_{\mathcal{E}_{co}}^{f,x} \setminus \Sigma_{\mathcal{E}_\sigma}^{f,x} \neq \emptyset$ and $\Sigma_{\mathcal{E}_\sigma}^{f,x} \setminus \Sigma_{\mathcal{E}_{co}}^{f,x} \neq \emptyset$ for $\sigma \in \{cf, ad\}$.

4.5 Remarks on Unrestricted AFs and Intertranslatability

Recently, some first results regarding expressibility w.r.t. unrestricted frameworks were presented in [Baumann and Spanring, 2017]. Remember that the set of unrestricted frameworks, abbreviated by \mathcal{F} , contains all AFs $F = (A, R)$, s.t. $A \subseteq \mathcal{U}$. This means, \mathcal{F} contains finite as well as infinite AFs and especially, AFs possessing all available arguments. It is obvious that signatures w.r.t. unrestricted frameworks contain more realizable sets than their finite counterparts since finite AFs may realize finite as well as finitely many extensions only. The following definition formally captures all considered types of signatures (cf. Definitions 4.4, 4.42 and 4.55) without any finite assumption.

Definition 4.61. Given a semantics σ . We call the set S the

1. *(unrestricted) σ -signature* if $S = \{\sigma(F) \mid F \in \mathcal{F}\}$ abbreviated by Σ_σ ,
2. *compact σ -signature* if $S = \{\sigma(F) \mid F \in CAF\}$ abbreviated by Σ_σ^c and
3. *analytic σ -signature* if $S = \{\sigma(F) \mid F \in XAF\}$ abbreviated by Σ_σ^x .

In [Baumann and Spanring, 2017] the authors were interested in a comparison of the expressive power of several mature semantics in the unrestricted setting. The following result shows that the relation between unrestricted signatures is intimately connected to their relation in case of finite, compact signatures. More precisely, non-empty relative complements in case of finite, compact signatures between two semantics carry over to their unrestricted versions. The main reason for this relation is the fact that unrestricted frameworks may contain any available argument of the universe \mathcal{U} .

Theorem 4.62 ([Baumann and Spanring, 2017]). *Given two semantics σ and τ with $\sigma, \tau \in \{na, stb, stg, ss, pr, co, gr, il, eg, cf, ad\}$ we have:*

1. *If $\Sigma_{\mathcal{E}_\sigma}^{f,c} \setminus \Sigma_{\mathcal{E}_\tau}^{f,c} \neq \emptyset$, then $\Sigma_{\mathcal{E}_\sigma}^c \setminus \Sigma_{\mathcal{E}_\tau}^c \neq \emptyset$ and*
2. *If $\Sigma_{\mathcal{E}_\sigma}^c \setminus \Sigma_{\mathcal{E}_\tau}^c \neq \emptyset$, then $\Sigma_{\mathcal{E}_\sigma} \setminus \Sigma_{\mathcal{E}_\tau} \neq \emptyset$.*

The following example illustrates the main proof idea.

Example 4.63. Let $\mathcal{E} \in \Sigma_{\mathcal{E}_{pr}}^{f,c} \setminus \Sigma_{\mathcal{E}_{stb}}^{f,c}$ (cf. Figure 4.3) and $F = (A, R)$ a witnessing framework. This means, F is finite, $\mathcal{E}_{pr}(F) = \mathcal{E}$ and pr -compact, i.e. $\bigcup \mathcal{E} = A$. Consider now $H = (\mathcal{U}, R)$. Obviously, $\mathcal{E}' = \mathcal{E}_{pr}(H) = \{E \cup (\mathcal{U} \setminus A) : E \in \mathcal{E}\}$ and $\bigcup \mathcal{E}' = \mathcal{U}$ showing the σ -compactness of H . In particular, $\mathcal{E}' \in \Sigma_{\mathcal{E}_{pr}}^c$. Note that any stb -realization of \mathcal{E}' has to be compact too since there are no additional arguments available. Assume $\mathcal{E}' \in \Sigma_{\mathcal{E}_{stb}}^c$, i.e. there is an AF $G' = (\mathcal{U}, R')$, s.t. $\mathcal{E}_{stb}(G') = \mathcal{E}'$. We observe that due to conflict-freeness there can not be attacks in G' between arguments from A and $\mathcal{U} \setminus A$ nor between any of the arguments from $\mathcal{U} \setminus A$. Consequently, $G = (A, R')$ is finite, $\mathcal{E}_{stb}(G) = \mathcal{E}$ and stb -compact implying that $\mathcal{E} \in \Sigma_{\mathcal{E}_{stb}}^{f,c}$ in contradiction to the initial assumption.

Now we are prepared for a comparison in case of unrestricted frameworks. Ignoring the superscripts in Figure 4.3 provides you with a graphical representation for selected semantics.

Theorem 4.64. *For unrestricted signatures the following hold:*

1. $\{\{E\} \mid E \subseteq \mathcal{U}\} = \Sigma_{\mathcal{E}_\sigma} \subset \Sigma_{\mathcal{E}_{na}} \subset \Sigma_{\mathcal{E}_\tau}$ for $\sigma \in \{gr, il\}$, $\tau \in \{stb, stg, ss, pr\}$,
2. $\Sigma_{\mathcal{E}_{eg}} \subset \Sigma_{\mathcal{E}_{pr}}$,
3. $\Sigma_{\mathcal{E}_{stb}} \subset \Sigma_{\mathcal{E}_\sigma}$ for $\sigma \in \{stg, ss\}$,

4. $\Sigma_{\mathcal{E}_{pr}} \setminus (\Sigma_{\mathcal{E}_{stb}} \cup \Sigma_{\mathcal{E}_{ss}} \cup \Sigma_{\mathcal{E}_{stg}}) \neq \emptyset$,
5. $\Sigma_{\mathcal{E}_{stg}} \setminus (\Sigma_{\mathcal{E}_{stb}} \cup \Sigma_{\mathcal{E}_{pr}} \cup \Sigma_{\mathcal{E}_{ss}}) \neq \emptyset$,
6. $\Sigma_{\mathcal{E}_{stb}} \setminus \Sigma_{\mathcal{E}_{pr}} \neq \emptyset$,
7. $\Sigma_{\mathcal{E}_{ss}} \setminus (\Sigma_{\mathcal{E}_{stb}} \cup \Sigma_{\mathcal{E}_{pr}} \cup \Sigma_{\mathcal{E}_{stg}}) \neq \emptyset$,
8. $\Sigma_{\mathcal{E}_{co}} \setminus \Sigma_{\mathcal{E}_{\sigma}} \neq \emptyset$ and $\Sigma_{\mathcal{E}_{\sigma}} \setminus \Sigma_{\mathcal{E}_{co}} \neq \emptyset$ for $\sigma \in \{cf, ad\}$,
9. $\Sigma_{\mathcal{E}_{cf}} \subseteq \Sigma_{\mathcal{E}_{ad}}$.

Finally, we briefly consider the closely related topic of *intertranslatability*. Intertranslatability revolves around the idea of mapping one semantics to another. One main motivation for studying this issue is the possibility to reuse a solver for one semantics for another [Dvořák and Woltran, 2011]. The main tool for this endeavour are functions mapping AFs to AFs, so-called *translations* formally defined as follows.

Definition 4.65. [Dvořák and Woltran, 2011] Given two semantics σ, τ . A function $f : \mathcal{F} \rightarrow \mathcal{F}$ is called an *exact translation* for $\sigma \rightarrow \tau$, if $\sigma(F) = \tau(f(F))$ for any AF F . It is called a *faithful translation* if for any AF F first $|\sigma(F)| = |\tau(f(F))|$ and second $\sigma(F) = \{S \cap A(F) \mid S \in \tau(f(F))\}$.

Please note that for some semantics there are no exact translations available due to reasons inherent to those semantics. For instance, preferred semantics satisfies *I-maximality*, i.e. for any AF F , $\mathcal{E}_{pr}(F)$ forms a \subseteq -antichain [Baroni and Giacomin, 2007b]. This implies that an exact translation $\mathcal{E}_{ad} \rightarrow \mathcal{E}_{pr}$ can not exist since for $F = (\{a\}, \emptyset)$ we observe $\{\emptyset, \{a\}\} = \mathcal{E}_{ad}(F)$. Sticking to faithful translations provides us with a positive answer if we consider finite AFs only [Spanring, 2012, Translation 3.1.85]. Interestingly, the considered translation does not serve in the general unrestricted case and interestingly, it was shown that a search for a suitable translation will never succeed (cf. [Baumann and Spanring, 2017, Example 6]).

The following theorem (a generalization of the finite version from [Dvořák and Spanring, 2017, Section 6.1]) establishes a close relation between realizability and intertranslatability as promised, namely: if τ is not less expressive than σ , then σ can be exactly translated to τ and vice versa.

Theorem 4.66 ([Baumann and Spanring, 2017]). *Given semantics σ, τ . We have: $\Sigma_{\sigma} \subseteq \Sigma_{\tau}$ if and only if there is an exact translation for $\sigma \rightarrow \tau$.*

The following Figure 4.6 illustrates translational (im)possibilities in an eye-catching way. Figure 4.6b summarizes known results regarding faithful translations in the finite case [Dvořák and Woltran, 2011; Spanring, 2012; Dvořák and Spanring, 2017], augmented with obvious insights for unique status semantics *il* and *eg*. For semantics σ, τ , encirclement in the same component indicates bidirectional translations. An arrow from σ to τ means directional translations. If there is no directed path (for instance for *na* to *cf*, or for *cf* to *gr*), then there is no translation. Figure 4.6a features the same visualization for unrestricted AFs. Dropping the finiteness restriction has some further consequences for the considered semantics, namely exact and faithful intertranslatability coincide. It is an open question whether both forms of translations are essentially the same in the general unrestricted setting. In consideration of Theorem 4.66 we may interpret Figure 4.6a as a comparison of the expressiveness of the considered semantics. That is, $\Sigma_{\sigma} \subseteq \Sigma_{\tau}$ if and only if there is a directed path from σ to τ .

As a final note, in contrast to the unrestricted setting Baumann and Spanring observed that for slightly restricted AFs $F = (A, R)$, s.t. $|A| \leq |\mathcal{U} \setminus A|$ it is possible to provide exact and efficiently computable translations from preferred to semi-stable semantics via $f(F) = F' = (A', R')$ with $A' = A \cup \{a' \mid a \in A\}$ and $R' = R \cup \{(a, a'), (a', a') \mid a \in A\}$. It is an interesting question whether this restriction allows for similar translational possibilities as in case of finite AFs.

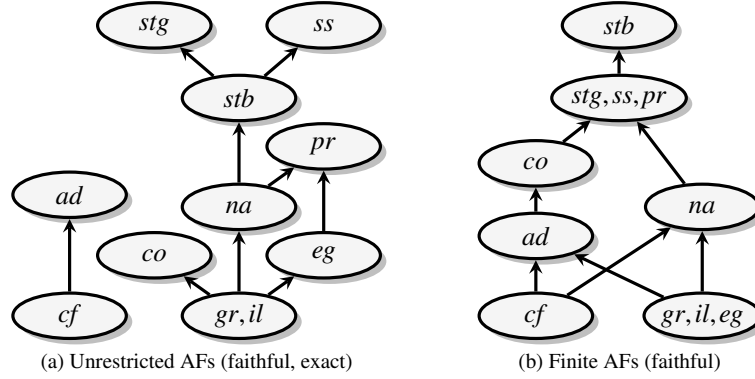


Figure 4.6: Translational (Im)Possibilities

4.6 Realizability and Signatures for Labelling-based Versions

Although any considered semantics σ possesses an extension-based version (indicated by \mathcal{E}_σ) as well as a closely related 3-valued labelling-based version (indicated by \mathcal{L}_σ) we formally have that both versions are different semantics (or more precisely, functions) in the sense of Definition 3.3. This formal difference has some impact on realizability as well as signatures. Let us consider realizability in the realm of finiteness. As a matter of fact, for any considered 3-valued labelling-based version \mathcal{L}_σ we have: if $F = (A, R)$ and $\mathcal{L} = (\mathcal{L}^i, \mathcal{L}^o, \mathcal{L}^u) \in \mathcal{L}_\sigma(F)$, then $A = \mathcal{L}^i \cup \mathcal{L}^o \cup \mathcal{L}^u$. This means, σ -labellings assign a status to any argument in F . Now, in case of finite AFs we know that potentially realizable sets of labellings have to involve finitely many arguments only. Moreover, these finitely many arguments precisely determine the set A of witnessing AFs $F = (A, R)$.⁶ Consider therefore the following example.

Example 4.67. Consider the following set of 3-valued labellings $\mathbb{S} = \{(\{a\}, \emptyset, \{b, c\}), (\{a, b\}, \{c\}, \emptyset)\}$. Is \mathbb{S} *co-realizable*? Since $\{a\} \cup \emptyset \cup \{b, c\} = \{a, b, c\}$ we deduce that candidates have to be members of $\mathcal{C} = \{F = (A, R) \mid A = \{a, b, c\}\}$. Note that $|\mathcal{C}| = 2^{|\{a, b, c\}|^2} = 2^9 = 512$. Clearly, this is a huge number, but it is a finite one. Consequently, the question of realizability can be decided by computing the σ -labellings of all AFs in \mathcal{C} . Of course, any intelligent search algorithm would involve further information like $\{a, b\}$ has to be conflict-free in a witnessing AF. Such an observation would decrease the number of candidates to $2^5 = 32$. However, in both cases one would find the unique witnessing framework F , i.e. $\mathcal{L}_{co}(F) = \mathbb{S}$, as depicted below.



The example above shows that the search space can be very large even in case of small numbers of arguments. Consequently, locally verifiable necessary as well as sufficient properties for realizability just like in case of extension-based semantics are of high interest too. To the best of our knowledge only two papers have dealt with labelling-based realizability in the context of AFs. The first study was presented by Dyrkolbotn [Dyrkolbotn, 2014]. The author showed that, as long as additional arguments are allowed any finite set of labellings is realizable under preferred and semi-stable semantics. It is important to emphasize that Dyrkolbotn uses a more relaxed notion of realizability, namely *realizability under projection* (cf. Definition 4.72). The other work [Linsbichler et al., 2016] deals with the standard

⁶This is exactly the point which does not carry over to finite realizability in case of extension-based semantics (cf. statement 2 of Theorem 4.44).

notion of finite realizability (Definition 4.3). The authors presented an algorithm which returns either “No” in case of non-realizability or a witnessing AF F in the positive case. The algorithm is not purely a guess-and-check method since it also includes a propagation step where certain necessary properties of witnessing AFs are processed. Remarkably, the algorithm is not restricted to the formalism of abstract argumentation frameworks only. In fact, it can also be used to decide realizability in case of the more general abstract dialectical frameworks as well as various of its sub-classes [Brewka and Woltran, 2010b; Brewka et al., 2013b].

4.6.1 Preliminary Results for Labelling-based Signatures

In the following we shed light on general relations between the labelling-based and extension-based signatures of the considered semantics. Fortunately, due to former characterization results we will even achieve characterizing or at least necessary properties for finite realizability regarding labelling-based versions. We proceed with the definition of an *labelling-set* which is the n -valued analogon (for $n \geq 2$) to an extension-set as introduced in Definition 4.5. A labelling-set is a finite set of n -tuples which are dealing with the same set of arguments and moreover, any n -tuple assigns exactly one status to each argument in question.

Definition 4.68. Given $S \subseteq (2^{\mathcal{U}})^n$. $Args_S$ denotes $\bigcup_{\mathcal{L}=(\mathcal{L}_1, \dots, \mathcal{L}_n) \in S} \bigcup_{i=1}^n \mathcal{L}_i$ and $\|S\|$ stands for $|Args_S|$. We say that S is a *labelling-set* if

1. $\|S\|$ is a finite cardinal,
2. for any $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_n) \in S$, $Args_S = \bigcup_{i=1}^n \mathcal{L}_i$ and
3. for any $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_n) \in S$, $\mathcal{L}_1, \dots, \mathcal{L}_n$ are pairwise disjoint.

The following proposition establishes a connection between extension-based and labelling-based realizability for any considered semantics. Roughly speaking, it states that labelling-based realizability requires extension-based realizability of the corresponding sets of in-labelled arguments. For a 3-tuple $\mathcal{L} = (\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ we also write $(\mathcal{L}^i, \mathcal{L}^o, \mathcal{L}^u)$ as usual.

Proposition 4.69. Given a set of 3-tuples $S \subseteq (2^{\mathcal{U}})^3$. For any semantics $\sigma \in \{stb, ss, stg, cf2, stg2, pr, ad, co, gr, il, eg, na, cf\}$ we have,

1. $S \in \Sigma_{\mathcal{L}_\sigma} \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\} \in \Sigma_{\mathcal{L}_\sigma}$ (unrestricted realizability)
2. $S \in \Sigma_{\mathcal{L}_\sigma}^c \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\} \in \Sigma_{\mathcal{L}_\sigma}^c$ (compact realizability)
3. $S \in \Sigma_{\mathcal{L}_\sigma}^x \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\} \in \Sigma_{\mathcal{L}_\sigma}^x$ (analytic realizability)
4. $S \in \Sigma_{\mathcal{L}_\sigma}^f \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\} \in \Sigma_{\mathcal{L}_\sigma}^f$ (finite realizability)
5. $S \in \Sigma_{\mathcal{L}_\sigma}^{f,c} \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\} \in \Sigma_{\mathcal{L}_\sigma}^{f,c}$ (finite, compact realizability)
6. $S \in \Sigma_{\mathcal{L}_\sigma}^{f,x} \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\} \in \Sigma_{\mathcal{L}_\sigma}^{f,x}$ (finite, analytic realizability)

Please note that the implications above are justified for any semantics σ whenever the different versions of it satisfy $\mathcal{E}_\sigma(F) = \{\mathcal{L}^i \mid \mathcal{L} \in \mathcal{L}_\sigma(F)\}$ for any relevant AF F . In the former sections we already presented characterization theorems or at least necessary properties for being finitely realizable regarding extension-based versions (cf. Theorems 4.12, 4.24 and 4.33). Combining these results with the proposition above yields the following necessary properties for finite realizability in the labelling-based case.

Note that the mentioned implications apply to finite, compact as well as finite, analytic signatures too since $\Sigma_{\mathcal{L}_\sigma}^{f,c} \subseteq \Sigma_{\mathcal{L}_\sigma}^f$ as well as $\Sigma_{\mathcal{L}_\sigma}^{f,x} \subseteq \Sigma_{\mathcal{L}_\sigma}^f$ by definition. In case of grounded, ideal and eager semantics we have that being an one-element labelling-set is necessary and even sufficient for being finitely realizable. One may easily verify that the only-if-directions of these semantics are justified by the witnessing framework $F_{\mathcal{L}} = (\mathcal{L}^i \cup \mathcal{L}^o \cup \mathcal{L}^u, \{(i, o) \mid i \in \mathcal{L}^i, o \in \mathcal{L}^o\} \cup \{(u, u) \mid u \in \mathcal{L}^u\})$ given that $S = \{\mathcal{L}\}$.

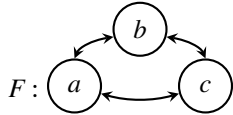
Theorem 4.70. *Given a set of 3-tuples $S \subseteq (2^{\mathcal{U}})^3$, then*

1. $S \in \Sigma_{\mathcal{L}_{cf}}^f \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\}$ is a non-empty, downward-closed and tight extension-set,
2. $S \in \Sigma_{\mathcal{L}_{na}}^f \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\}$ is a non-empty, incomparable extension-set and $dcl(S)$ is tight,
3. $S \in \Sigma_{\mathcal{L}_{gr}}^f \iff S$ is a labelling-set with $|S| = 1$,
4. $S \in \Sigma_{\mathcal{L}_{il}}^f \iff S$ is a labelling-set with $|S| = 1$,
5. $S \in \Sigma_{\mathcal{L}_{eg}}^f \iff S$ is a labelling-set with $|S| = 1$,
6. $S \in \Sigma_{\mathcal{L}_{stb}}^f \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\}$ is a incomparable and tight extension-set,
7. $S \in \Sigma_{\mathcal{L}_{sig}}^f \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\}$ is a non-empty, incomparable and tight extension-set,
8. $S \in \Sigma_{\mathcal{L}_{ad}}^f \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\}$ is a conflict-sensitive extension set containing \emptyset ,
9. $S \in \Sigma_{\mathcal{L}_{pr}}^f \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\}$ is a non-empty, incomparable and conflict-sensitive extension-set,
10. $S \in \Sigma_{\mathcal{L}_{ss}}^f \implies \{\mathcal{L}^i \mid \mathcal{L} \in S\}$ is a non-empty, incomparable and conflict-sensitive extension-set.

4.6.2 Realizability under Projection

We turn now to *realizability under projection* which was firstly considered in [Dyrkolbotn, 2014]. In order to realize a set of labellings S under projection it suffices to come up with an AF F , s.t. its set of labellings restricted to the relevant arguments coincide with S . Consider therefore the following illustrating example.

Example 4.71. Given $S = \{(\{a\}, \{b\}, \emptyset), (\{b\}, \{a\}, \emptyset), (\emptyset, \{a, b\}, \emptyset)\}$. We observe that the corresponding set of sets of in-labelled arguments $S^I = \{\emptyset, \{a\}, \{b\}\}$ violates incomparability. Thus, applying statement 9 of Theorem 4.70 we derive that S is not finitely *pr*-realizable. Consider now the following AF F .



We obtain $\mathcal{L}_{pr}(F) = \{(\{a\}, \{b, c\}, \emptyset), (\{b\}, \{a, c\}, \emptyset), (\{c\}, \{a, b\}, \emptyset)\}$. Now, if we restrict any labelling $\mathcal{L} = (\mathcal{L}^i, \mathcal{L}^o, \mathcal{L}^u) \in \mathcal{L}_{pr}(F)$ to the arguments a and b , i.e. $\mathcal{L}|_{\{a, b\}} = (\mathcal{L}^i \cap \{a, b\}, \mathcal{L}^o \cap \{a, b\}, \mathcal{L}^u \cap \{a, b\})$ we obtain exactly all labellings in S . In this sense, S is *pr*-realizable under projection.

We proceed with the formal definitions. For the sake of completeness we introduce realizability under projection and its corresponding signatures w.r.t. any kind of semantics as defined in Definition 3.3.

Definition 4.72. Given a semantics $\sigma : \mathcal{F} \rightarrow 2^{(2^{\mathcal{U}})^n}$. A set $S \subseteq (2^{\mathcal{U}})^n$ is σ -realizable under projection if there is an AF F , s.t. $\sigma(F)|_{\text{Args}_S} = \{E|_{\text{Args}_S} \mid E \in \mathcal{E}_\sigma(F)\} = S$ (in case of $n = 1$) or $\sigma(F)|_{\text{Args}_S} = \{\mathcal{L}|_{\text{Args}_S} \mid \mathcal{L} \in \mathcal{L}_\sigma(F)\} = S$ (for $n \geq 2$), respectively.

Definition 4.73. Given a semantics σ . The unrestricted as well as finite σ -projection-signatures are defined as follows:

1. $\Sigma_\sigma^p = \{\sigma(F)|_B \mid F = (A, R) \in \mathcal{F}, B \subseteq A\}$ and
2. $\Sigma_\sigma^{f,p} = \{\sigma(F)|_B \mid F = (A, R) \in \mathcal{F}, F \text{ is finite}, B \subseteq A\}$

Analogously to Proposition 4.69 we state the following relation between labelling-based and extension-based versions of the considered semantics.

Proposition 4.74. Given a set of 3-tuples $S \subseteq (2^{\mathcal{U}})^3$. For any semantics $\sigma \in \{stb, ss, stg, cf2, stg2, pr, ad, co, gr, il, eg, na, cf\}$ we have,

1. $S \in \Sigma_{\mathcal{L}_\sigma}^p \implies \{\mathcal{L}' \mid \mathcal{L} \in S\} \in \Sigma_{\mathcal{E}_\sigma}^p$ (unrestricted realizability under projection)
2. $S \in \Sigma_{\mathcal{L}_\sigma}^{f,p} \implies \{\mathcal{L}' \mid \mathcal{L} \in S\} \in \Sigma_{\mathcal{E}_\sigma}^{f,p}$ (finite realizability under projection)

As a matter of fact, any projection signature is a superset of the corresponding signature. The following question then arises naturally: how much more sets can be generated if we stick to realizability under projection? For instance, we have already seen that even comparable sets are realizable under projection by semantics satisfying incomparability (Example 4.71). It was the main result in [Dyrkolbotn, 2014, Theorem 3.1] that in case of semi-stable and preferred semantics indeed any 3-valued labelling-set is finitely realizable under projection. The proof relies on two basic constructions. The first step *generates* an AF, consisting of so-called *circuits*, s.t. its set of preferred as well as semi-stable labellings restricted to the relevant arguments contains any possible labelling. The second construction *eliminates* undesired labellings step by step. Combining this realizability result with statement 2 of Proposition 4.74 yields the following theorem.

Theorem 4.75. Let $\sigma \in \{pr, ss\}$. We have,

1. $\Sigma_{\mathcal{L}_\sigma}^{f,p} = \{S \subseteq (2^{\mathcal{U}})^3 \mid S \text{ is a labelling-set}\}$ and
2. $\Sigma_{\mathcal{E}_\sigma}^{f,p} = \{S \subseteq 2^{\mathcal{U}} \mid S \text{ is an extension-set}\}.$

4.7 Final Remarks and Conclusion

We have dealt with different forms of realizability in the context of abstract argumentation frameworks. In accordance with the existing literature the main part of this section was devoted to finite realizability for extension-based semantics. However, for any semantics σ we may state the following general subset relations depicted as Venn-diagram.

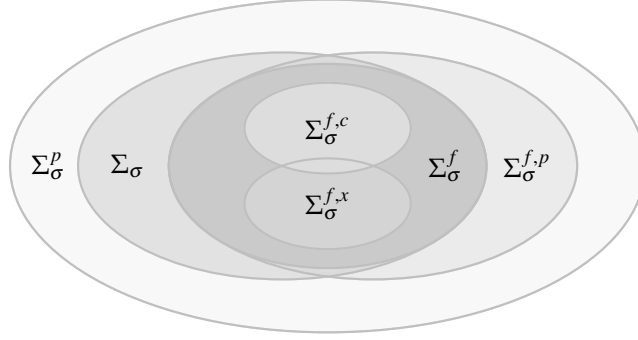


Figure 4.7: Subset Relations between Different Kinds of Signatures

In case of the extension-based versions of naive, grounded, ideal, eager, stable, stage, preferred and semi-stable semantics as well as conflict-free and admissible sets we provided exact characterizations for their corresponding general signatures. We have seen that for some semantics we do not lose any expressive power if sticking to compact or analytic AFs, i.e. $\Sigma_\sigma^f = \Sigma_\sigma^{f,c}$ or $\Sigma_\sigma^f = \Sigma_\sigma^{f,x}$, respectively. However, for certain prominent semantics, e.g. preferred semantics we have that the expressive power indeed relies on the use of rejected arguments or implicit conflicts. For such semantics, it remains an open problem to present exact characterizations for finite, compact or finite, analytic realizability, respectively. In case of labelling-based versions of semi-stable and preferred semantics we have seen that any labelling-set is realizable under projection. In [Dyrkolbotn, 2014] it was already noted that this equality does not hold for any semantics. For instance, the empty labelling is admissible for any AF F . Hence, in case of admissible semantics, no labelling-set is realizable under projection if it fails to include the empty labelling.

Finally, let us mention some computational issues not considered so far. It can be said that on the one hand, the classes of finite, compact and finite, analytic provide computational benefits both in practice and in terms of theoretical worst-case analysis. On the other hand testing for membership in one of the classes is, for most of the semantics, of rather high complexity and thus these classes cannot be directly used to improve systems. We refer the interested reader to [Baumann et al., 2016a] for more details. Moreover, in general, given an extension-set S , deciding whether S is compactly realizable is a hard problem, that is, by definition of the decision problem there are no good reasons to believe that we can do any better than guessing a compact AF and checking whether its extension-set coincides with S . Nevertheless, for some semantics we have seen that finite, compact realizability can be characterized locally, i.e. by properties of S itself (as shown in Theorem 4.45). In this case, finite, compact realizability can be checked in polynomial time as for standard finite realizability [Dunne et al., 2015, Theorem 6]. Moreover, in [Baumann et al., 2016a] a huge number of shortcuts to detect non-compactness are provided. By shortcut we mean a property of the given extension-set S that is easily computable (preferably in polynomial time) which (sometimes) provides us with a definitive answer to the decision problem. These shortcuts are related to numerical aspects of argumentation frameworks like results concerning maximal number of extensions [Baumann and Strass, 2013].

Chapter 5

Replaceability of Subframeworks

Given a certain logical formalism \mathcal{L} and two syntactically different \mathcal{L} -theories T_1 and T_2 . One central question is whether, and if so, how to decide that these \mathcal{L} -theories represent the same information? Of course, in order to answer this question we have to clarify what we exactly mean by sharing the same information first. Note that there is neither a uniquely determined, nor a certain preferred interpretation by the formalism \mathcal{L} itself. For instance, equating information with possessing the same semantics yields to the well-known notion of *ordinary* or *standard equivalence*. This means, assuming that $\sigma_{\mathcal{L}}$ is the semantics of \mathcal{L} we might answer that T_1 and T_2 are equivalent if and only if $\sigma_{\mathcal{L}}(T_1) = \sigma_{\mathcal{L}}(T_2)$. A more demanding interpretation of sharing the same information is to require that T_1 and T_2 are semantically indistinguishable even if further \mathcal{L} -theories T are added to both simultaneously. More formally, we may state: T_1 and T_2 are considered to be equivalent if and only if $\sigma_{\mathcal{L}}(T_1 \cup T) = \sigma_{\mathcal{L}}(T_2 \cup T)$ for any theory T . This notion is known as *strong equivalence* and is of high interest for any logical formalism since it allows one to locally replace, and thus give rise for simplification, parts of a given theory without changing the semantics of the latter. In contrast to classical logics where standard and strong equivalence coincide (cf. Chapter 2), it is possible to find ordinary but not strongly equivalent objects for any nonmonotonic formalism available in the literature. Consequently, much effort has been devoted to characterizing strong equivalence for nonmonotonic formalisms such as logic programs [Lifschitz et al., 2001], causal theories [Turner, 2004], default logic [Turner, 2001] as well as nonmonotonic logics in general [Truszczyński, 2006].

In [Oikarinen and Woltran, 2011] the authors introduced the notion of strong equivalence for abstract AFs. They provided a series of characterization theorems for deciding strong equivalence of two AFs with respect to several semantics. In view of the fact that strong equivalence is defined semantically it is the main and quite surprisingly insight that being strongly equivalent can be decided syntactically. More precisely, they introduced the notion of a *kernel* of an AF F which is (informally speaking) a subgraph of F where certain attacks are deleted and showed that syntactical identity of suitably chosen kernels characterizes strong equivalence w.r.t. the considered semantics. Strong equivalence is, as its name suggests, a very (and often unnecessarily to) strong notion of equivalence if dynamic evolvments are considered. In many argumentation scenarios the type of modification which may potentially occur can be anticipated and furthermore, more importantly, does not range over *arbitrary expansions* as required for strong equivalence. Let us consider the instantiation-based context where AFs are built from an underlying knowledge base. Here, we typically observe that older arguments and their corresponding attacks survive and only new arguments which may interact with the previous ones arise given that a new piece of information is added to the underlying knowledge base. This type of dynamic evolvment is a so-called *normal expansion* and its corresponding equivalence notion were firstly studied in [Baumann, 2012a]. Over the last five years several equivalence notions taking into account specific types of evolvments reflecting

the nature of various argumentation scenarios were defined and characterized. The considered dynamic scenarios range from the most general form, so-called *updates* [Baumann, 2014a] where arguments and attacks can be deleted and added to different types of *expansions* [Oikarinen and Woltran, 2011; Baumann, 2012a; Baumann and Brewka, 2010] and *deletions* [Baumann, 2014a] where arguments and/or attacks are allowed to be added or deleted in a certain way only.

Into the year 2015 all characterization theorems were stated in terms of extension-based semantics. Recently, Baumann presents their labelling-based counterparts and showed that, although labelling-based semantics contain more information than their extension-based counterpart, there is a majority of equivalence relations where labelling-based and extension-based versions coincide [Baumann, 2016]. Even more recently, a first consideration of strong equivalence regarding unrestricted frameworks were presented in [Baumann and Spanring, 2017]. It turned out that there are no characterizational differences compared to the finite case as long as the AFs in question are *jointly expandable*, i.e. that the existence of fresh arguments is guaranteed.

Another approach somehow complementary to the ones mentioned before is presented in [Baroni et al., 2014] where sharing the same information is interpreted as possessing the same Input/Output behavior. Roughly speaking, the main idea is to consider an argumentation framework as a kind of black box which receives some input from the external world (i.e. a set of external arguments) via incoming attacks and produces an output to the external world via outgoing attacks. Such an interacting module is called an *argumentation multipole*. Two multipoles connected with the same external world are considered as *Input/Output equivalent* if the effects, i.e. the produced labellings for external arguments are the same for any reasonable input-labelling. This notion yields the possibility of replacing a multipole with another one embedded in a larger framework without affecting the labellings of the unmodified part of the initial framework. The interested reader is referred to [Baroni et al., 2018b] for further information. In the following we shed light on equivalence notions induced by certain dynamic scenarios.

5.1 Dynamic Scenarios and Corresponding Equivalence Notions

There are two main classes of dynamic scenarios, namely *expansions* and *deletions*. Both of them can be further divided in *normal* and *local* versions. These scenarios are motivated by real-world argumentation as well as instantiation-based argumentation [Caminada and Amgoud, 2007]. For instance, let us consider the dynamics of a discussion or dispute illustrated by the following citation [Besnard and Hunter, 2009]:

How does argumentation usually take place? Argumentation starts when an initial argument is put forward, making some claim. An objection is raised, in the form of a counterargument. The latter is addressed in turn, eventually giving rise to a counter-counterargument, if any. And so on.

This means, in order to strengthen the own point of view or to rebut the opponents arguments it is natural that one tries to come up with *stronger* arguments, i.e. new arguments which are not attacked by the former arguments. This type of dynamics is formally captured by so-called *strong expansions* [Baumann and Brewka, 2010]. The formal counterpart of it, so-called *weak expansions* [Baumann and Brewka, 2010], where the new arguments do not attack (but may be attacked by) the old ones seem to be more an academic exercise than a task with practical relevance with regard to real-world argumentation.¹ Let us turn to instantiation-based argumentation where arguments and attacks stem from an underlying knowledge base (cf. Chapter 6 for detailed information as well as Figure 2 in Chapter 4 for an illustration). What happens on the abstract level if a new piece of information is added? It turns out that in almost all deductive argumentation systems older arguments and their corresponding attacks survive and only

¹We mention that they do play a decisive role w.r.t. computational issues, so-called *splitting methods* (cf. [Baumann, 2011; Baumann et al., 2011; Baumann et al., 2012]).

new arguments which may interact with the previous ones arise. This type of dynamic evolvement is formally captured by so-called *normal expansions*. *Local expansions* in contrast, i.e. expansions where new attacks are added only correspond to re-instantiations if we change to a less restrictive notion of attack (cf. [Besnard and Hunter, 2001] for different attack notions).

We start with the definition of the different types of expansions together with some introducing examples.

Definition 5.1 ([Baumann and Brewka, 2010]). An AF G is an *expansion* of AF $F = (A, R)$ (for short, $F \preceq_E G$) iff $G = (A \cup B, R \cup S)$ for some (maybe empty) sets B and S , s.t. $A \cap B = R \cap S = \emptyset$. An expansion is called

1. *normal* ($F \preceq_N G$) iff $\forall ab ((a, b) \in S \rightarrow a \in B \vee b \in B)$,
2. *strong* ($F \preceq_S G$) iff $F \preceq_N G$ and $\forall ab ((a, b) \in S \rightarrow \neg(a \in A \wedge b \in B))$,
3. *weak* ($F \preceq_W G$) iff $F \preceq_N G$ and $\forall ab ((a, b) \in S \rightarrow \neg(a \in B \wedge b \in A))$,
4. *local* ($F \preceq_L G$) iff $B = \emptyset$.

For short, being a normal expansion means that new attacks must involve at least one new argument in contrast to local expansions where new attacks involve old arguments only. Moreover, strong and weak expansions are normal and their names refer to properties of the additional arguments, namely arguments which are never attacked by former arguments (so-called *strong* arguments) and arguments which do not attack former arguments (so-called *weak* arguments).

Observe that any arbitrary expansion can be splitted up in a normal and a local part. This can be nicely seen in the following example.

Example 5.2. The AF F is the initial framework. An arbitrary, normal, strong, weak or local expansion of it are given by F_E , F_N , F_S , F_W and F_L , respectively. Grey-highlighted arguments or attacks represent added information.

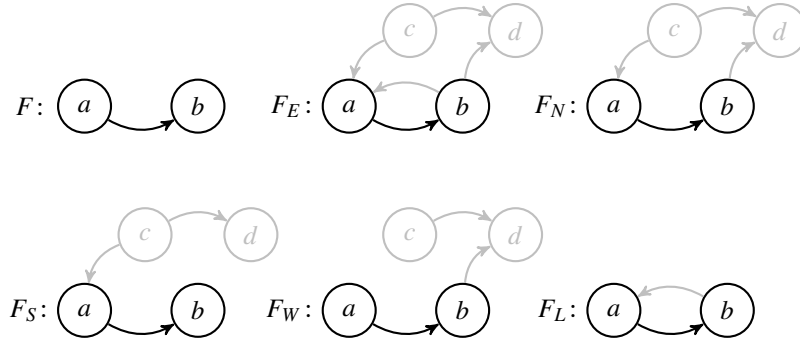


Figure 5.1: Different Kinds of Expansions

In 2014 the natural counter-parts (or more precisely, inverse operations) to arbitrary, normal and local expansions, so-called *deletions* were introduced [Baumann, 2014a]. Furthermore, the most general form of a dynamic scenario (where expansion and deletion can be combined) a so-called *update* were considered too. Analogously to expansions, any arbitrary deletion can be splitted in a normal and a local part. This means, a *normal deletion* retract arguments and their corresponding attacks. *Local deletions* in

contrast delete attacks only.² The main motivation behind these notions stems from instantiation-based context. More precisely, a normal deletion on the abstract level correspond to deleting information of a given knowledge base. Changing to a more restrictive notion of attack correspond to a local deletion and a combination of both of them give rise to an arbitrary deletion on the abstract level. We proceed with the formal definitions as well as introductory examples.

Definition 5.3 ([Baumann, 2014a]). Given an AF $F = (A, R)$, a set of arguments B and a set of attacks S as well as a further AF H . The AF

$$G = (F \setminus [B, S]) \cup H = ((A, R \setminus S)|_{A \setminus B}) \cup H$$

is called an *update* of F (for short, $F \preceq_U G$). An update is called a

1. *deletion* ($F \succeq_D G$) iff $H = (\emptyset, \emptyset)$,
2. *normal deletion* ($F \succeq_{ND} G$) iff ($F \succeq_D G$) and $S = \emptyset$,
3. *local deletion* ($F \succeq_{LD} G$) iff $F \succeq_D G$ and $B = \emptyset$.

Let us take a closer look at the definition of $G = (F \setminus [B, S]) \cup H$. The AF H plays the role of added information, i.e. it contains new arguments and attacks. Consequently, for all kind of deletions we have $H = (\emptyset, \emptyset)$ which leaves us with $G = F \setminus [B, S]$. The set B contains arguments which have to be deleted. Since attacks depend on arguments we have to delete the attacks which involve arguments from B too. This operation is formally captured by the restriction of F to $A \setminus B$. Furthermore, the set S contains particular attacks which have to be deleted. This means, the pair $[B, S]$ does not necessarily have to be an AF. Therefore we use $[B, S]$ instead of (B, S) . If clear from context we use B and S instead of $[B, \emptyset]$ or $[\emptyset, S]$, i.e. we simply write $F \setminus B$ as well as $F \setminus S$ for normal or local deletions, respectively.

Example 5.4. The AF F represents the initial situation. An update as well as arbitrary, normal or local deletion of it are given by F_U , F_D , F_{ND} and F_{LD} . Grey-highlighted arguments or attacks represent added information in contrast to dotted arguments and attacks which represent deleted objects.³ More formally, in accordance with Definition 5.3 we have that $F_U = (F \setminus [B, S]) \cup H$, $F_D = F \setminus [B, S]$, $F_{ND} = F \setminus B$, $F_{LD} = F \setminus S$ where the set of arguments $B = \{c\}$, the set of attacks $S = \{(b, a)\}$ and the AF $H = (\{b, d, e, f\}, \{(d, b), (e, f), (f, d)\})$.

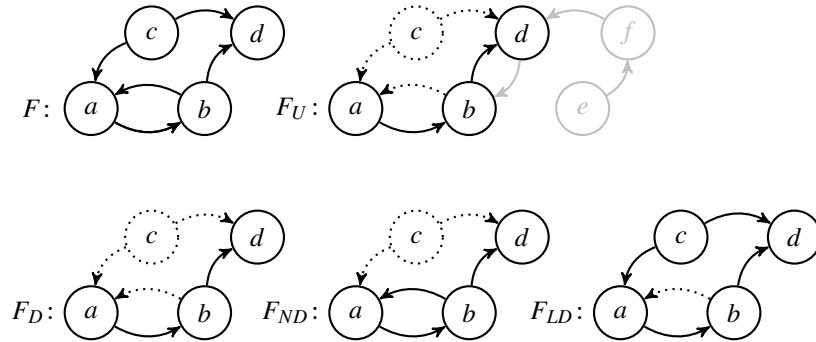


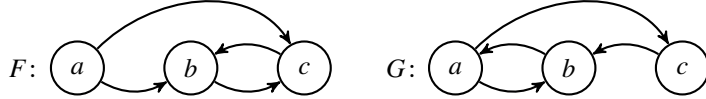
Figure 5.2: An Update and Different Kinds of Deletions

²We mention that *strong* as well as *weak deletions* are not introduced/considered so far. They could be easily defined as inverse operations of their expansion counterparts. Before doing so, it would be interesting to identify real-world situations or instantiation-based dynamics where such kind of evolutions naturally occur.

³This convention will be used throughout the whole chapter.

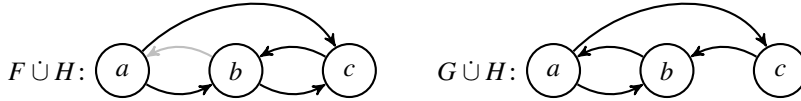
We now turn to the corresponding equivalence notions (cf. [Baumann and Strass, 2015, Section 3.8] for chronological order). Two AFs F and G are said to be *ordinarily equivalent* w.r.t. a semantics σ if they possess the same σ -extensions/labellings. In this case, we say that F and G possess the same *explicit* information. In contrast, sharing the same *implicit* information, i.e. being semantically indistinguishable w.r.t. any suitable future scenario is a much more demanding property which allows to replace F and G by each other without loss of semantical information.

Example 5.5. Consider the following AFs F and G . We have $\mathcal{E}_{pr}(F) = \mathcal{E}_{pr}(G) = \{\{a\}\}$. This means, F and G possess the same explicit information w.r.t. preferred semantics or in other words, they are ordinarily equivalent.

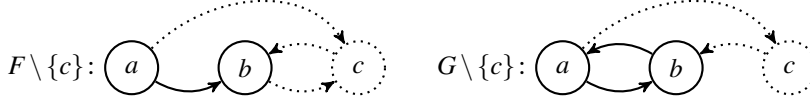


Assume that expansions as well deletions are the dynamic scenarios of interest. This means, we ask whether the AFs F and G even possess the same implicit information w.r.t. expansions or deletions, respectively? In order to give a negative answer one has to come up with one single dynamic scenario where the revised versions possess different preferred extensions. A positive answer in contrast is a statement about infinitely many dynamic scenarios (even in case of finite AFs). In this example, we give a negative answer for both modification types.

In case of expansions, we conjoin to both the AF $H = (\{a, b\}, \{(b, a)\})$. Consider the resulting frameworks below. We have $\mathcal{E}_{pr}(F \cup H) = \{\{a\}, \{b\}\}$ and since $G \cup H = G$ we obtain $\mathcal{E}_{pr}(G \cup H) = \{\{a\}\}$ without re-computing.



To reveal the inherent difference between F and G in case of deletions we may retract with the argument c . Consider the resulting (normal) deletions $F \setminus \{c\}$ and $G \setminus \{c\}$ of F or G , respectively. Now, $\{b\}$ becomes a preferred extension in $F \setminus \{c\}$ but still not in $G \setminus \{c\}$.



We now formally define what we precisely mean by possessing the same implicit information. As already stated, the first paper in this line of work was [Oikarinen and Woltran, 2011] engaged with characterizing *strong equivalence*. For the sake of clarity and comprehensibility we use the term *expansion equivalence* since strong equivalence [Oikarinen and Woltran, 2011, Definition 2] corresponds to semantical indistinguishability w.r.t. arbitrary expansions.

Before presenting the definitions let us introduce some further notational convention which will be extensively used throughout the whole chapter. For a given AF $G = (B, S)$, we use $A(G) = B$, $R(G) = S$, $L(G) = \{a \in B \mid (a, a) \in S\}$ and $NL(G) = B \setminus L(G)$.

Definition 5.6. Given a semantics σ . Two AFs F and G are

1. *ordinarily equivalent* w.r.t. σ (for short, $F \equiv^\sigma G$) iff $\sigma(F) = \sigma(G)$,
2. *expansion equivalent* w.r.t. σ (for short, $F \equiv_E^\sigma G$) iff for each AF H we have, $F \cup H \equiv^\sigma G \cup H$,
3. *normal expansion equivalent* w.r.t. σ (for short, $F \equiv_N^\sigma G$) iff for each AF H , such that $F \preceq_N F \cup H$ and $G \preceq_N G \cup H$ we have, $F \cup H \equiv^\sigma G \cup H$,

4. *strong expansion equivalent w.r.t. σ* (for short, $F \equiv_S^\sigma G$) iff for each AF H , such that $F \preceq_S F \dot{\cup} H$ and $G \preceq_S G \dot{\cup} H$ we have, $F \dot{\cup} H \equiv^\sigma G \dot{\cup} H$,
5. *weak expansion equivalent w.r.t. σ* (for short, $F \equiv_W^\sigma G$) iff for each AF H , such that $F \preceq_W F \dot{\cup} H$ and $G \preceq_W G \dot{\cup} H$ we have, $F \dot{\cup} H \equiv^\sigma G \dot{\cup} H$,
6. *local expansion equivalent⁴ w.r.t. σ* (for short, $F \equiv_L^\sigma G$) iff for each AF H with $A(H) \subseteq A(F \dot{\cup} G)$ we have, $F \dot{\cup} H \equiv^\sigma G \dot{\cup} H$.
7. *update equivalent w.r.t. σ* (for short, $F \equiv_U^\sigma G$) iff for any pair $[B, S]$ and any AF H we have, $(F \setminus [B, S]) \dot{\cup} H \equiv^\sigma (G \setminus [B, S]) \dot{\cup} H$,
8. *deletion equivalent w.r.t. σ* (for short, $F \equiv_D^\sigma G$) iff for any pair $[B, S]$ we have, $F \setminus [B, S] \equiv^\sigma G \setminus [B, S]$,
9. *normal deletion equivalent w.r.t. σ* (for short, $F \equiv_{ND}^\sigma G$) iff for any set of arguments B we have, $F \setminus B \equiv^\sigma G \setminus B$,
10. *local deletion equivalent w.r.t. σ* (for short, $F \equiv_{LD}^\sigma G$) iff for any set of attacks S we have, $F \setminus S \equiv^\sigma G \setminus S$,

Remember that there are several relations between the considered dynamic scenarios. For instance, in accordance with Definitions 5.1 and 5.3, any normal expansion (deletion) is an arbitrary expansion (deletion). Furthermore, in the light of Definition 5.6, we certainly affirm that expansion equivalence is much more demanding than local expansion equivalence. In other words, local expansion equivalence of two AFs is an immediate and unavoidable consequence of being expansion equivalent. Finally, any considered equivalence notion is at least as demanding than ordinary equivalence.⁵ Please note that these relations do not depend on certain properties of a considered semantics. Consequently, Figure 5.3 gives a preliminary overview for such interrelations (arising from the definitions) between the introduced equivalence notions for any possible semantics. For reasons, which will become clearer later, we also consider the identity relation. For two equivalence notion Φ and Ψ we have $\Phi \subseteq \Psi$ iff there is a link from Φ to Ψ .

In the remainder of this section we shed light on the question of *how* to determine whether two AFs are equivalent w.r.t. certain scenarios? As a by-product of these characterization results we will see that for many semantics the preliminary relations between the introduced equivalence notions depicted above can be delineated in a much more compact way. The majority of the presented characterization results is devoted to finite AFs as well as extension-based semantics. We will see that there are some differences if sticking to unrestricted frameworks or the corresponding labelling-based versions.

5.2 Characterization Theorems for Extension-based Semantics

5.2.1 The Central Notion of Expansion Equivalence

In order to get an idea of how to find a characterization we start with some reflections. For this purpose we consider the most restrictive semantics, namely the stable one as well as the most prominent type of equivalence, namely expansion equivalence. What are necessary features of expansion equivalence w.r.t. stable semantics, i.e. which properties are implied if two AFs F and G are expansion equivalent? In consideration of Figure 5.3 we deduce their ordinary equivalence, i.e. $\mathcal{E}_{stb}(F) = \mathcal{E}_{stb}(G)$. Note that

⁴Note that a suitable AF H is not necessarily a local expansion of F and G in the sense of Definition 5.1. Nevertheless, we may loosely speak about local expansions.

⁵The empty framework (\emptyset, \emptyset) as well as the empty pair $[\emptyset, \emptyset]$ justifies this assertion for any type of expansions or deletions, respectively.

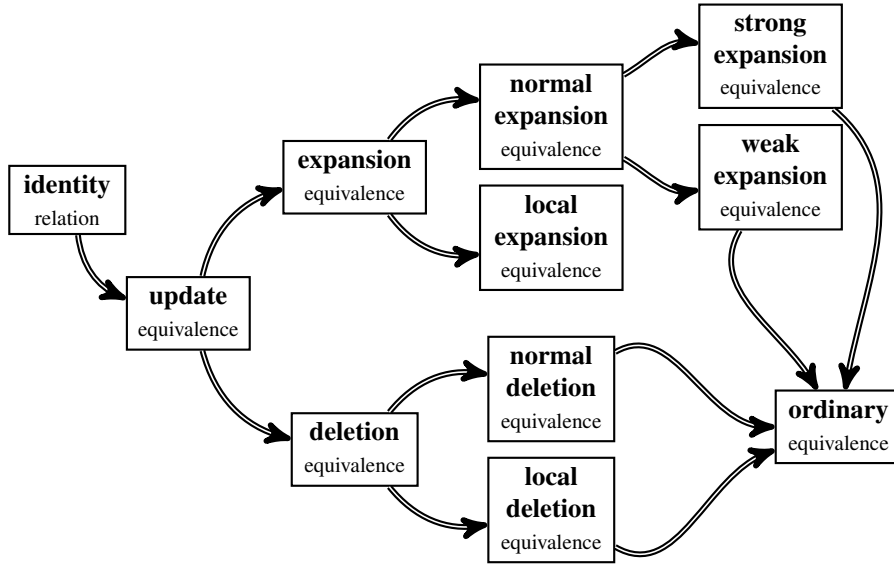
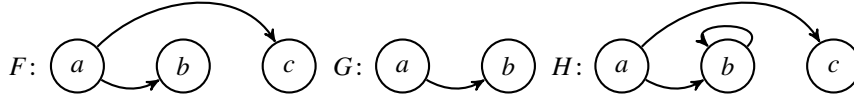


Figure 5.3: Preliminary Subset Relations between Equivalence Notions

possessing the same set of extensions neither imply sharing the same arguments nor sharing the same self-loops as shown in the following example.

Example 5.7. Consider the AFs F , G and H . Each two of them are ordinarily equivalent since $\mathcal{E}_{stb}(F) = \mathcal{E}_{stb}(G) = \mathcal{E}_{stb}(H) = \{\{a\}\}$.



The AFs $I_1 = (\{c\}, \emptyset)$ and $I_2 = (\{a, b, c\}, \{(b, a), (b, c)\})$ witness that neither F and G , nor F and H are expansion equivalent w.r.t. stable semantics. Convince yourself that $\mathcal{E}_{stb}(F \cup I_1) = \{\{a\}\} \neq \{\{a, c\}\} = \mathcal{E}_{stb}(G \cup I_1)$ and $\mathcal{E}_{stb}(F \cup I_2) = \{\{a\}, \{b\}\} \neq \{\{a\}\} = \mathcal{E}_{stb}(H \cup I_2)$.

Restricting ourselves to finite AFs, it is not difficult to see that in case of expansion equivalence w.r.t. stable semantics the observed relation between non-sharing the same arguments/loops and non-equivalence does hold in general. In other words, possessing the same arguments as well as possessing the same loops are indeed necessary conditions for being expansion equivalent in the finite setting.

Let us summarize our observations in the following fact.

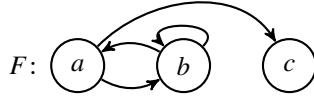
Fact 5.8. Given two finite AFs F and G . If $F \equiv_E^{\mathcal{E}_{stb}} G$, then

1. $\mathcal{E}_{stb}(F) = \mathcal{E}_{stb}(G)$,
2. $A(F) = A(G)$ and
3. $L(F) = L(G)$.

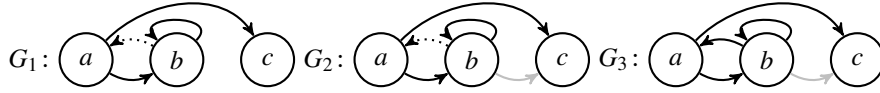
As already stated in Figure 5.3, being identical (i.e. $A(F) = A(G)$ and $R(F) = R(G)$) is sufficient for being expansion equivalent. Combining this undeniable fact together with the second and third items of Fact 5.8 encourages one to search for syntactical properties sufficient as well as necessary for being

expansion equivalent. In order to guarantee the first item of Fact 5.8 we have to identify attacks which do not contribute anything when computing stable extensions. Moreover, these attacks which do not affect the evaluation of a given AF F have to be *redundant*, no matter how F is extended. Remember that being a stable extension can be simply verified by checking whether the set in question is conflict-free and possesses a full range.⁶ This means, good candidates for “useless” attacks w.r.t. stable semantics should fulfill the following two properties: firstly, having or not having such an attack does not change the status of a set from being conflict-free to conflicting or vice versa and secondly, having or not having such an attack does not affect the range of a conflict-free set. Certainly, an attack (a, b) stemming from a self-defeating argument a does not change the conflict status of a certain set E . This can be seen as follows: If $a \in E$, then E was conflicting as well as remains conflicting after deleting or adding (a, b) . Furthermore, if $a \notin E$, then E might be conflicting or not. In either case the conflict status of E does not change if (a, b) is added or removed since $\{a, b\} \not\subseteq E$. Finally, such an attack (a, b) might have an influence on the range of conflicting sets but it definitely has not in case of conflict-free sets since $a \notin E$ can not be questioned.

Example 5.9. Consider the following AF F . We have, $\mathcal{E}_{stb}(F) = \{\{a\}\}$.



According to our considerations above adding or deleting an attack stemming from the self-defeating argument b does not change the semantics. Consider therefore the following three possible “manipulations”.



Indeed, $\mathcal{E}_{stb}(F) = \mathcal{E}_{stb}(G_1) = \mathcal{E}_{stb}(G_2) = \mathcal{E}_{stb}(G_3) = \{\{a\}\}$ support our claims for the static case. We encourage the reader to try to do the impossible, namely semantically distinguish the AFs F and its manipulations by an arbitrary expansion.

It was the main result in [Oikarinen and Woltran, 2011] that expansion equivalence can be indeed decided by looking at the syntax only. The authors introduced so-called *kernels* which are simply functions mapping each AF F to its redundancy-free version. This means, the kernel of an AF F does not possess any redundant attack. Put it differently, for any surviving attack exist at least one dynamic scenario were deleting this attack would cause a semantical difference. We proceed with the formal definition of the very first kernels already introduced in [Oikarinen and Woltran, 2011]. We sometimes call them *classical*.

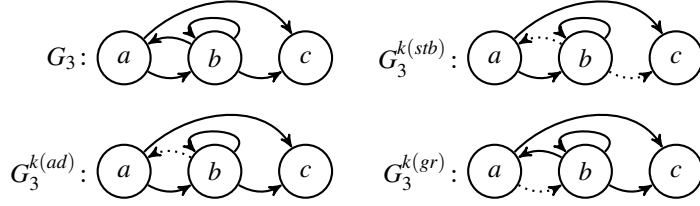
Definition 5.10. Let $\sigma \in \{stb, ad, gr, co\}$. The σ -kernel $k(\sigma) : \mathcal{F} \rightarrow \mathcal{F}$ with $k(\sigma)(F) = F^{k(\sigma)} = (A, R^{k(\sigma)})$ for a given AF $F = (A, R)$ is defined as:

$$\begin{aligned} R^{k(stb)} &= R \setminus \{(a, b) \mid a \neq b, (a, a) \in R\}, \\ R^{k(ad)} &= R \setminus \{(a, b) \mid a \neq b, (a, a) \in R, \{(b, a), (b, b)\} \cap R \neq \emptyset\}, \\ R^{k(gr)} &= R \setminus \{(a, b) \mid a \neq b, (b, b) \in R, \{(a, a), (b, a)\} \cap R \neq \emptyset\}, \\ R^{k(co)} &= R \setminus \{(a, b) \mid a \neq b, (a, a), (b, b) \in R\}. \end{aligned}$$

In order to get an idea of how the classical kernels work we proceed with an example.

⁶The topic of *verifiability* of argumentation semantics σ will be considered in Chapter 6. The main question is which (minimal amount of) information on top of conflict-free sets is exactly needed to determine whether a certain set is a σ -extension.

Example 5.11. Consider again the AF G_3 depicted in Example 5.9. We apply now all classical kernels.



The stable kernel deletes all attacks (a, b) stemming from a self-defeating argument a . A deletion of (a, b) in case of the grounded kernel additionally requires that a is counter-attacked by b or b is self-defeating or both. Interchanging a and b yields the condition for deletion in case of the grounded kernel. Finally, $G_3^{k(co)} = G_3$ since deleting an attack (a, b) w.r.t. the complete kernel requires that both arguments a and b are self-defeating.

Before turning to characterization theorems, we collect some useful properties of the introduced kernels. The following fact contains intrinsic properties of the classical kernels.⁷ More precisely, any classical kernel k is *node-preserving* and *loop-preserving*, i.e. the sets of arguments and self-defeating arguments do not change if applying k . Moreover, in the absence of self-loops, each AF coincides with its classical kernels. Furthermore, the decision whether an attack (a, b) has to be deleted does not depend on further arguments than a and b . Put differently, the reason of being redundant is *context-free*, i.e. it stems from the arguments themselves. The last two properties claim that equality of kernels is *robust* w.r.t. further compositions as well as deleting arguments and corresponding attacks. For a given AF $F = (B, S)$ we use $A(F)$, $R(F)$ and $L(F)$ to refer to its arguments, attacks and self-defeating arguments, i.e. $A(F) = B$, $R(F) = S$ and $L(F) = \{a \in A(F) \mid (a, a) \in R(F)\}$.

Fact 5.12 (cf. [Oikarinen and Woltran, 2011; Baumann, 2014a]). *Given $k \in \{k(stb), k(ad), k(gr), k(co)\}$. For any finite AF F we have:*

1. $A(F) = A(F^k)$, (node-preserving)
2. $L(F) = L(F^k)$, (loop-preserving)
3. $L(F) = \emptyset \implies F = F^k$ and (sufficient condition for identity)
4. $(a, b) \in R(F^k) \iff (a, b) \in R\left(\left(F|_{\{a, b\}}\right)^k\right)$. (context-freeness)

Furthermore, for finite AFs F and G we have:

5. If $F^k = G^k$, then $(F \dot{\cup} H)^k = (G \dot{\cup} H)^k$ for any finite AF H and ($\dot{\cup}$ -robustness)
6. If $F^k = G^k$, then $(F \setminus B)^k = (G \setminus B)^k$ for any finite set of args B . (\setminus -robustness)

We proceed with extrinsic properties, i.e. features of kernels in presence of semantics. More precisely, stable, admissible, grounded and complete semantics are insensitive w.r.t. the application of their corresponding classical σ -kernel, i.e. the set of σ -extensions remains unchanged. Furthermore, the admissible kernel neither effects semi-stable, eager, preferred and ideal semantics. Similarly in case of stable kernel and stage semantics.

⁷Although most of the properties are immediately clear even in case of unrestricted frameworks we will state all of them for finite AFs only as done in the existing literature. The same applies to Fact 5.18. Some results regarding unrestricted frameworks can be found in Section 5.3.

Fact 5.13 ([Oikarinen and Woltran, 2011; Gaggl and Woltran, 2013]). *For any finite AF F we have:*

1. $\mathcal{E}_\sigma(F) = \mathcal{E}_\sigma(F^{k(\sigma)})$ for $\sigma \in \{stb, ad, gr, co\}$,
2. $\mathcal{E}_\sigma(F) = \mathcal{E}_\sigma(F^{k(ad)})$ for $\sigma \in \{ss, eg, pr, il\}$ and
3. $\mathcal{E}_{stg}(F) = \mathcal{E}_{stg}(F^{k(stb)})$.

As already mentioned, kernels play a decisive role in deciding expansion equivalence. In general, we say that an equivalence notion \equiv is *characterizable through k* or simply, *k is a characterizing kernel (of \equiv)* if for any two AFs F and G , $F \equiv G$ iff $F^k = G^k$. This means, proving whether two frameworks are equivalent can be done by simply checking whether the corresponding kernels are identical. Note that all classical kernels can be efficiently constructed from a given AF. The following main theorem states that for all nine considered semantics σ there is a certain classical kernel k , s.t. expansion equivalence w.r.t. σ is characterizable through k in the finite setting. This is a very remarkable result since expansion equivalence is defined semantically. For instance, two finite AFs F and G are expansion equivalent w.r.t. stable semantics if and only if the associated stable kernels $F^{k(stb)}$ and $G^{k(stb)}$ are syntactically equal. Observe that there is no need to introduce further kernels since one single kernel may serve for different semantics.

Theorem 5.14. [Oikarinen and Woltran, 2011; Gaggl and Woltran, 2013] *For finite AFs F, G we have:*

1. $F \equiv_E^\sigma G \iff F^{k(\sigma)} = G^{k(\sigma)}$ for any $\sigma \in \{stb, ad, co, gr\}$,
2. $F \equiv_E^\sigma G \iff F^{k(ad)} = G^{k(ad)}$ for any $\sigma \in \{pr, il, ss, eg\}$ and
3. $F \equiv_E^{stg} G \iff F^{k(stb)} = G^{k(stb)}$.

Having Theorem 5.14 at hand we can now formally verify that all AFs depicted in Example 5.9 are expansion equivalent w.r.t. stable semantics. This means, the recommended search for arbitrary expansions revealing semantical difference between them will never succeed. As an aside, one might get the impression that the syntactical characterization presented in Theorem 5.14 is somehow unique. This is not true. Consider therefore the equivalence class $[F]_E^{stb} = \{G \mid F \equiv_E^{stb} G\}$ induced by F . Mathematically speaking, the stable kernel $F^{k(stb)}$ represents the least (w.r.t. subgraph-relation) element in $[F]_E^{stb}$. It is not difficult to prove that $[F]_E^{stb}$ even possesses a greatest element, namely $F^{k'(stb)} = (A(F), R(F) \cup \{(a, b) \mid a \neq b, (a, a) \in R(F)\})$, i.e. the framework resulting from F by adding (instead of deleting) all redundant attacks. In case of finite AFs it can be shown with reasonable effort that expansion equivalence w.r.t. stable semantics is characterizable through $k'(stb)$ too. In the same manner, all other semantics considered in Theorem 5.14 possess alternative “greatest elements” characterizations. We will see that the so-called *naive kernel* (compare Definition 5.17) provides such a kind of characterization for naive semantics. The reason for this “choice” is simply that the induced equivalence classes do not necessarily possess a least element in case of naive semantics.

Finally, let us turn to the more exotic cf2 as well as stage2 semantics which are defined via a recursive schema based on the decomposition of AFs along their strongly connected components (SCCs). These semantics are exceptional regarding expansion equivalence since in contrast to all other semantics considered in this section we have that even attacks between two self-attacking arguments are *meaningful*. This means, the presence or absence of such attacks may change the outcome of an AF. Moreover, it turned out that any attack is non-redundant. In summary, expansion equivalence coincides with syntactical identity or more formally, for any finite AF F , $|[F]_E^{cf2}| = |[F]_E^{stg2}| = |\{F\}| = 1$.

Theorem 5.15. [Gaggl and Woltran, 2013; Gaggl and Dvořák, 2016] Given $\sigma \in \{cf2, stg2\}$. For finite AFs F and G we have,

$$F \equiv_E^{\sigma} G \iff F = G.$$

5.2.2 Further Equivalence Notions Characterizable through Kernels

Let us turn to the remaining equivalence notions? Are there similar syntax-based characterization results?

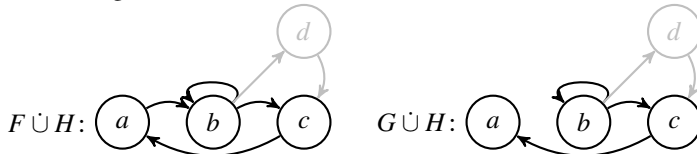
5.2.2.1 Weaker Notions of Expansion Equivalence

Let us consider less demanding notions than expansion equivalence, e.g. normal and local expansion equivalence. In consideration of Definition 5.1 we do not have good reasons to believe that two AFs could be semantically distinguished by normal or local expansions, given that we only have a witnessing arbitrary expansion showing their non-equivalence. It was one surprising result in this line of research, that for many semantics, expansion equivalence coincide with definitorially weaker notions of it. This implies that weaker notions than expansion equivalence can be characterized by classical kernels too. The first results in this respect were already given in [Oikarinen and Woltran, 2011, Theorem 8]. The authors showed that for some semantics expansion equivalence and local expansion equivalence coincide if considering finite AFs. It is worthwhile to gain a thorough understanding of this relation since it actually means that if there is an arbitrary expansion which semantically distinguish two finite AFs, then there has to be a local expansion doing likewise. Later it was shown that even normal expansion equivalence coincides with expansion equivalence for a whole bunch of semantics [Baumann, 2012a]. Interestingly, in contrast to local expansion equivalence, there are (to the best of our knowledge) no semantics together with witnessing AFs known which show that this coincidence does not hold in general.

Example 5.16. Consider the following AFs F and G . According to Theorem 5.14 they are not expansion equivalent w.r.t. preferred semantics since $F^{k(ad)} = F \neq G = G^{k(ad)}$.



As already stated (up to now) normal expansion equivalence coincides with expansion equivalence for any considered semantics. One possible scenario which makes the predicted different behaviour explicit is the following.



Formally, we define $H = (\{b, c, d\}, \{(b, d), (d, c)\})$ and we obtain $\{\{a, d\}\} = \mathcal{E}_{pr}(F \cup H) \neq \{\emptyset\} = \mathcal{E}_{pr}(G \cup H)$. We encourage the reader to try to find a witnessing example showing that F and G are not local expansion equivalent w.r.t. preferred semantics. Due to Theorem 5.20 there has to be at least one distinguishing local expansion.

How do the semantics behave in case of strong expansion equivalence? Remember, a special feature of strong expansions is that a former attack between old arguments will never become a counterattack to an added attack. In this sense, former attacks do not play a role with respect to being a potential defender of an added argument. Hence, in contrast to arbitrary expansions where such attacks might be relevant, we may delete them without changing the behavior with respect to further evaluations. To make this point clearer consider again the AF $F \cup H$ depicted in Example 5.16. Note that the already existing attack (a, b) in F becomes a defending attack of the newly added argument d . This means, such attacks

in fact play an important role with respect to further evaluation in case of arbitrary expansions. It was one main result in [Baumann, 2012a] that for some semantics attacks like (a, b) in F are indeed redundant w.r.t. strong expansions. Even more surprising, strong expansion equivalence is characterizable through kernels. Therefore, more involved kernel definitions, so-called σ -*-kernels had to be introduced. These kernels allow more deletions than their classical counterparts for expansion equivalence. In contrast to them, σ -*-kernels are *context-sensitive*, i.e. the question whether an attack (a, b) is redundant can not be answered by considering the arguments a and b only [Baumann, 2014a].

The first three kernels presented in the definition below were firstly introduced in [Baumann, 2012a] with the objective to characterize strong expansion equivalence with respect to certain semantics. For the sake of completeness we also present the so-called *stg*-*-kernel as well as *na*-kernel [Baumann and Woltran, 2016; Baumann et al., 2016c].⁸

Definition 5.17. Let $\sigma \in \{ad, gr, co, stg\}$. The σ -*-kernel $k^*(\sigma) : \mathcal{F} \rightarrow \mathcal{F}$ with $k^*(\sigma)(F) = F^{k^*(\sigma)} = (A, R^{k^*(\sigma)})$ for a given AF $F = (A, R)$ is defined as:

$$\begin{aligned} R^{k^*(ad)} &= R \setminus \{(a, b) \mid a \neq b, ((a, a) \in R \wedge \{(b, a), (b, b)\} \cap R \neq \emptyset) \\ &\quad \vee ((b, b) \in R \wedge \forall c ((b, c) \in R \rightarrow \{(a, c), (c, a), (c, c), (c, b)\} \cap R \neq \emptyset))\}, \\ R^{k^*(gr)} &= R \setminus \{(a, b) \mid a \neq b, ((b, b) \in R \wedge \{(a, a), (b, a)\} \cap R \neq \emptyset) \\ &\quad \vee ((b, b) \in R \wedge \forall c ((b, c) \in R \rightarrow \{(a, c), (c, a), (c, c)\} \cap R \neq \emptyset))\}, \\ R^{k^*(co)} &= R \setminus \{(a, b) \mid a \neq b, ((a, a), (b, b) \in R) \vee ((b, b) \in R \wedge (b, a) \notin R \\ &\quad \wedge \forall c ((b, c) \in R \rightarrow \{(a, c), (c, a), (c, c), (c, b)\} \cap R \neq \emptyset))\}, \\ R^{k^*(stg)} &= R \setminus \{(a, b) \mid a \neq b, (a, a) \in R \vee \forall c (c \neq a \rightarrow (c, c) \in R)\} \\ R^{k^*(na)} &= R \cup \{(a, b) \mid a \neq b, \{(a, a), (b, a), (b, b)\} \cap R \neq \emptyset\}. \end{aligned}$$

The latter represents the so-called *na*-kernel $F^{k(na)} = (A, R^{k(na)})$.

For an illustrating example we refer the reader to Example 5.19. Analogously to Fact 5.12 we collect some properties of the newly introduced kernels. The first three properties are immediately clear by definition.⁹ The robustness w.r.t. deletions and corresponding attacks is less obvious but it is already shown for all considered kernels (except the *stg*-*-kernel) in case of finite AFs (cf. [Baumann, 2014a, Theorems 6 and 14]).

Fact 5.18. Given $k \in \{k^*(ad), k^*(gr), k^*(co), k^*(stg), k(na)\}$ and $k^* \in \{k^*(ad), k^*(gr), k^*(co), k^*(stg)\}$. For two finite AFs F and G we have:

1. $A(F) = A(F^k)$, (node-preserving)
2. $L(F) = L(F^k)$, (loop-preserving)
3. $L(F) = \emptyset \implies F = F^{k^*}$ and (sufficient condition for identity)
4. If $F^k = G^k$, then $(F \setminus B)^k = (G \setminus B)^k$ for any finite set of arguments B . (\setminus -robustness)

⁸As an aside, we use the supplement “*”, whenever the kernel in question is non-classical and expansion equivalence is already characterized by another kernel.

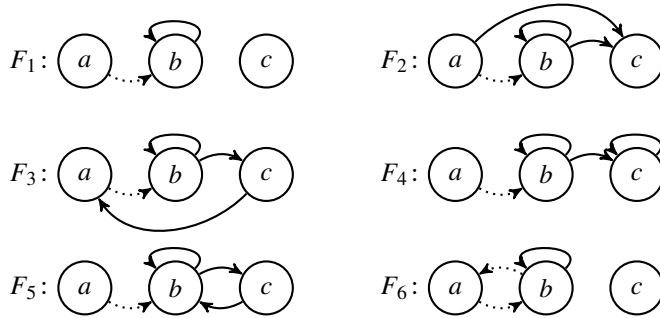
⁹The AF $F = (\{a, b\}, \{(a, b)\})$ shows that the naive kernel has to be excluded from item 3 of Fact 5.18 since $F^{k(na)} = (\{a, b\}, \{(a, b), (b, a)\}) \neq F$.

Let us consider the ad -*-kernel (which, as we shall see, characterizes strong expansion equivalence for preferred semantics) in more detail. Consider the first disjunct. This first condition is exactly the same as in case of the ad -kernel (compare Definition 5.10), i.e. an attack (a, b) has to be deleted if a is self-attacking and at least one of the attacks (b, a) or (b, b) exist. The second disjunct provides one with further options to delete an attack (a, b) , namely if b is self-defeating and furthermore, for all arguments c which are attacked by b at least one of the following conditions has to be fulfilled:

1. a attacks c ,
2. c attacks a ,
3. c attacks c ,
4. c attacks b .

The motivation for the second disjunct is the following: At first observe that b cannot be an element of any conflict-free set. Consequently, in case of strong expansions the attack (a, b) may only be relevant with respect to the defense of c . In the first three cases this relevance becomes unimportant since $\{a, c\}$ is conflicting. In the fourth case the redundancy of (a, b) with respect to the defense of c is given by the fact that c already defends itself against b . Please note that the consideration of $c = a$ or $c = b$ is not excluded by Definition 5.17. The following frameworks exemplify different cases.

Example 5.19. The following graphs show six frameworks and their corresponding ad -*-kernels. The dotted attacks represent initial attacks which have to be deleted if applying the ad -*-kernel.



Consider the formal description of $R^{k^*(ad)}$ as given in Definition 5.17. The AF F_1 is somehow the base case since the only argument c , s.t. $(b, c) \in R(F_1)$ is b itself. Since $(b, b) \in R(F_1)$ we deduce that the considered intersection is non-empty and thus, the deletion of (a, b) is justified. The subsequent four frameworks F_2, F_3, F_4 and F_5 are the base case plus one further argument c different from a and b , s.t. for any $i \in \{2, 3, 4, 5\}$, $(b, c) \in R(F_i)$. The last framework F_6 illustrates the case b counterattacks a . Note that the reason to delete (a, b) is somehow self-referential since (additionally to the base case) it is justified by $(a, b) \in R(F_6)$. Due to the first disjunct (i.e. just like in case of the classical ad -kernel) even the attack (b, a) has to be deleted.

We proceed with further characterization theorems.¹⁰ An comprehensive overview of equivalence notion and their characterizing kernels in case of finite AFs and extension-based semantics is presented in Figure 5.4.

¹⁰Please note that the results in case of cf2 and stage2 semantics have never been published before.

Theorem 5.20. [Oikarinen and Woltran, 2011; Baumann, 2012a; Baumann and Woltran, 2016; Baumann et al., 2016b] For finite AFs F and G we have the following coincidences.

1. $F \equiv_E^{\sigma} G \iff F \equiv_N^{\sigma} G$ for $\sigma \in \{stg, stb, ss, eg, ad, pr, il, gr, co, na, cf2, stg2\}$,
2. $F \equiv_E^{\sigma} G \iff F \equiv_L^{\sigma} G$ for $\sigma \in \{ss, eg, ad, pr, il, na\}$ and
3. $F \equiv_E^{\sigma} G \iff F \equiv_S^{\sigma} G$ for $\sigma \in \{stg, stb, ss, eg, na\}$.

Furthermore, for any two finite AFs F and G we have the following non-classical characterizations.

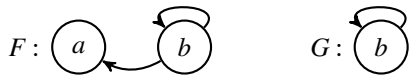
4. $F \equiv_L^{stg} G \iff F^{k^*(stg)} = G^{k^*(stg)}$,
5. $F \equiv_S^{eg} G \iff F^{k^*(ad)} = G^{k^*(ad)}$ for $\sigma \in \{ad, pr, il\}$,
6. $F \equiv_S^{\sigma} G \iff F^{k^*(\sigma)} = G^{k^*(\sigma)}$ for $\sigma \in \{co, gr\}$ and
7. $F \equiv_E^{na} G \iff F^{k(na)} = G^{k(na)}$.

At this point we want to highlight a very surprising relation. Remember that normal expansion equivalence and normal deletion equivalence are completely unrelated in the general picture (cf. Figure 5.3). The observation that the characterizing kernels (including the identity map in case of cf2 and stage2 semantics) of normal expansion equivalence w.r.t. all considered semantics in this section satisfy \backslash -robustness (cf. Facts 5.12 and 5.18) reveals that normal expansion equivalence implies normal deletion equivalence for these semantics.

Corollary 5.21. Given $\sigma \in \{stg, stb, ss, eg, ad, pr, il, gr, co, na, cf2, stg2\}$ and two finite AFs F and G . We have: $F \equiv_N^{\sigma} G \implies F \equiv_{ND}^{\sigma} G$.

The attentive reader may have noticed that we do not have characterized local expansion equivalence w.r.t. stable, complete as well as grounded extension-based semantics. We mention that all three equivalence notions are already characterized but the characterization theorems are not purely kernel-based (cf. [Oikarinen and Woltran, 2011, Theorems 9,10,11]). Furthermore, it can be checked that none of the kernels presented in Definitions 5.10 and 5.17 serve as a characterizing kernel. Consider therefore the following example [Oikarinen and Woltran, 2011, Example 15].

Example 5.22. The AFs F and G are local expansion equivalent w.r.t. stable semantics. This can be seen as follows. Given an AF H , s.t. $A(H) \subseteq \{a, b\}$. If $(a, b) \in R(H)$ and $(a, a) \notin R(H)$, we obtain $\mathcal{E}_{stb}(F \dot{\cup} H) = \mathcal{E}_{stb}(G \dot{\cup} H) = \{\{a\}\}$. Otherwise, $\mathcal{E}_{stb}(F \dot{\cup} H) = \mathcal{E}_{stb}(G \dot{\cup} H) = \emptyset$.



Remember that all introduced kernels are node-preserving (Facts 5.12 and 5.18). Consequently, none of them may serve as a characterizing kernel for local expansion equivalence w.r.t. stable semantics.

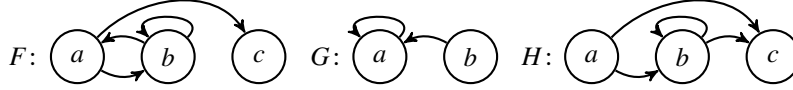
We mention that weak expansion equivalence is already characterized in case of stable semantics [Baumann, 2012a, Proposition 3] as well as admissible, preferred and complete semantics [Baumann and Brewka, 2018, Theorem 1]. All characterization results are not kernel-based. For instance, two AFs are weak expansion equivalent w.r.t. stable semantics iff both do not possess stable extensions at all or if they share the same arguments and at the same time possess the same stable extensions. Consequently, $F = (\{a\}, \{(a, a)\})$ and $G = (\{a, b, c\}, \{(a, b), (b, c), (c, a)\})$ are weak expansion equivalent w.r.t. stable semantics. Both frameworks witness that any potential characterizing kernel k is necessarily neither node- nor loop-preserving.

As a final note, we are not aware of any study of weaker notions of expansion equivalence in case of cf2 as well as stage2 semantics.

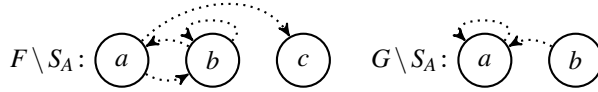
5.2.2.2 Notions of Deletion Equivalence and Update Equivalence

We start with local deletion equivalence. Remember that local deletion equivalent AFs cannot be semantically distinguished by deleting a certain set of attacks in both simultaneously. How “strong” is this notion? Are there redundant attacks or even redundant arguments?

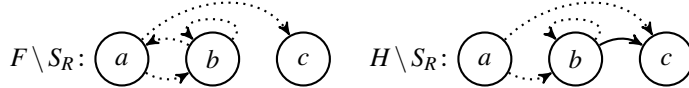
Example 5.23. Consider the following AFs F , G and H .



The AFs F and G do not possess the same arguments. Let us delete all occurring attacks, i.e. $S_A = R(F) \cup R(G)$. We obtain the following local deletions where $\{a, b, c\} \in \mathcal{E}_\sigma(F \setminus S_A) \setminus \mathcal{E}_\sigma(G \setminus S_A)$ for all semantics σ considered in this section.



The AFs F and H possess the same arguments but differ in their attack-relation, e.g. $(b, c) \in R(H) \setminus R(F)$. This difference can be made more explicit if defining $S_R = (R(F) \cup R(H)) \setminus \{(b, c)\}$. Consider the resulting local deletions.



Once again we have $\{a, b, c\} \in \mathcal{E}_\sigma(F \setminus S_R)$ for all known semantics σ and $\{a, b, c\} \notin \mathcal{E}_\sigma(H \setminus S_R)$ if assuming conflict-freeness of the considered semantics.

The observations above indicate that there is not much space for redundancy in case of local expansion equivalence and indeed, it was one main result in [Baumann, 2014a] that local expansion equivalence collapse to identity for all semantics considered in this section. Moreover, instead of proving this one by one for any semantics the author followed the line in [Baroni and Giacomin, 2007b] and provide abstract criteria guaranteeing the coincidence with syntactical identity. These criteria are very weak requirements, namely *conflict-freeness* (\mathcal{CF}) and the principle of *isolate-inclusion* (\mathcal{II}). The latter is fulfilled by a semantics σ iff for any AF F , the set of all isolated arguments is contained in at least one σ -extension. Observe that any considered semantics apart from stable semantics satisfy \mathcal{II} .¹¹

Theorem 5.24 ([Baumann, 2014a]). *Given a semantics σ satisfying \mathcal{CF} and \mathcal{II} . For two finite AFs F and G we have:*

$$F \equiv_{LD}^\sigma G \iff F = G.$$

Since being identical implies local deletion equivalence we deduce that all equivalence notion “in-between” them collapse to identity too (cf. Figure 5.3).

Proposition 5.25. *Given a semantics σ satisfying \mathcal{CF} and \mathcal{II} . For two finite AFs F and G we have:*

$$F \equiv_U^\sigma G \iff F \equiv_D^\sigma G \iff F = G.$$

This means, for semantics satisfying conflict-freeness and isolate-inclusion any argument/attack may play a crucial role with respect to further evaluations if updates, deletions or local deletions are considered. Note that the results may apply to future semantics. In order to refine the general picture (as depicted in Figure 5.3) for the semantics considered in this section we state the following relations.¹²

¹¹A counter-example in case of stable semantics is given by $F = (\{a, b\}, \{(b, b)\})$. Obviously, a is isolated but $\mathcal{E}_{stb}(F) = \emptyset$. Nevertheless, local expansion equivalence in case of stable semantics collapse to identity too.

¹²The results in case of cf2 and stage2 semantics have never been published before. It can be checked that both semantics satisfy the preconditions of Theorem 5.25.

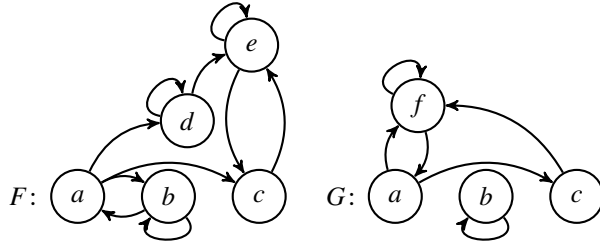
Corollary 5.26. Let $\sigma \in \{stg, stb, ss, eg, ad, pr, il, gr, co, na, cf2, stg2\}$. For two finite AFs F, G we have:

1. $F \equiv_U^{\mathcal{E}^\sigma} G \iff F \equiv_D^{\mathcal{E}^\sigma} G \iff F \equiv_{LD}^{\mathcal{E}^\sigma} G \iff F = G$, ($k = id$)
2. $F \equiv_D^{\mathcal{E}^\sigma} G \implies F \equiv_E^{\mathcal{E}^\sigma} G$, (deletion vs. expansion)
3. $F \equiv_{LD}^{\mathcal{E}^\sigma} G \implies F \equiv_L^{\mathcal{E}^\sigma} G$. (local versions)

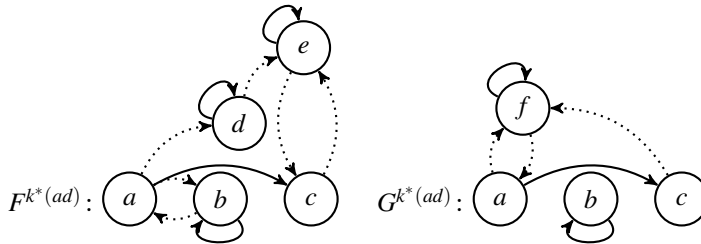
5.2.3 The Exceptional Case of Normal Deletion Equivalence

Normal deletion equivalence, where the retraction of arguments and corresponding attacks is considered, is exceptional in several regards. Firstly, the characterization theorems for admissible, complete and grounded semantics partially rely on σ -*-kernels. Remember that these kernels were originally introduced to characterize strong expansion (cf. Theorem 5.20). Secondly, normal deletion equivalent AFs do not even have to share the same arguments and thus give space for simplifications.

Example 5.27. Consider the following AFs F and G . We have $\mathcal{E}_{ad}(F) = \mathcal{E}_{ad}(G) = \{\emptyset, \{a\}\}$. Even more, for any set of arguments B , $\mathcal{E}_{ad}(F \setminus B) = \mathcal{E}_{ad}(G \setminus B)$ showing their normal deletion equivalence, i.e. $F \equiv_{ND}^{\mathcal{E}_{ad}} G$.



Observe that the non-shared arguments d, e and f do not play a role for the evaluation w.r.t. admissible semantics since firstly, they are self-defeating and thus cannot be part of an admissible set; and secondly, if they attack a non-looping argument shared by both arguments, e.g. e attacks c in F or f attacks a in G , then they are counter-attacked by the same argument, i.e. c attacks e in F and a attacks f in G . Consequently, they cannot influence potential admissible sets being a subset of $\{a, b\}$. Finally, let us consider the ad -*-kernel of both frameworks (cf. Example 5.19 and the comments above for more details).



Obviously, F and G do not possess the same kernels but note that their restrictions to the shared arguments do, i.e. $(F|_{\{a,b,c\}})^{k*(ad)} = (G|_{\{a,b,c\}})^{k*(ad)}$.

It turned out that the issues raised in Example 5.27 are essential to characterize normal deletion equivalence w.r.t. admissible semantics. In case of complete and grounded semantics slightly different conditions have to be fulfilled, namely w.r.t. the non-shared arguments we have “it is forbidden to

be attacked” instead of “counter-attack if attacked” like in case of admissible semantics and furthermore, instead of the ad -*-kernel the corresponding σ -*-kernels are used. Consider therefore the following definition and the characterization theorem. We use Δ to denote the symmetric difference, i.e. $A\Delta A' = A \setminus A' \cup A' \setminus A$. Moreover, $NL(F) = A(F) \setminus L(F)$, i.e. $NL(F)$ contains all arguments of F which are not self-defeating.

Definition 5.28. Given $F = (A, R)$ and $G = (A', R')$ and let $\sigma \in \{co, gr\}$.

1. $Loop(F, G) \iff_{def} L(F \dot{\cup} G|_{A\Delta A'}) = A\Delta A'$, (“non-shared args are self-defeating”)
2. $Att^{ad}(F, G) \iff_{def} \forall b \in A \setminus A' \forall a \in NL(F|_{A \cap A'}) : ((b, a) \in R \rightarrow (a, b) \in R) \wedge \forall b \in A' \setminus A \forall a \in NL(G|_{A \cap A'}) : ((b, a) \in R' \rightarrow (a, b) \in R')$, (“counter-attack if attacked”)
3. $Att^\sigma(F, G) \iff_{def} \forall b \in A \setminus A' \forall a \in NL(F|_{A \cap A'}) : (b, a) \notin R \wedge \forall b \in A' \setminus A \forall a \in NL(G|_{A \cap A'}) : (b, a) \notin R'$. (“it is forbidden to be attacked”)

Theorem 5.29 ([Baumann, 2014a]). Let $\sigma \in \{ad, co, gr\}$. Given two finite AFs $F = (A, R)$ and $G = (A', R')$ and let $I = A \cap A'$,

$$F \equiv_{ND}^\sigma G \iff Loop(F, G), Att^\sigma(F, G), (F|_I)^{k^*(\sigma)} = (G|_I)^{k^*(\sigma)}.$$

In contrast to admissible, complete and grounded semantics where normal deletion equivalence is indeed weaker than normal expansion equivalence we observe that these notions coincide in case of stable semantics. This means, normal deletion equivalence w.r.t. stable semantics is characterized by the classical stable kernel too.

The following theorem corrects the corresponding result in [Baumann, 2014a, Theorem 10] which did not take into account that an empty framework possess a stable extension, namely the empty one.¹³

Theorem 5.30. For finite AFs F and G we have:

$$F \equiv_{ND}^{\mathcal{E}_{stb}} G \iff F^{k(stb)} = G^{k(stb)}.$$

Proof. (\implies) We show the contrapositive, i.e. $F^{k(stb)} \neq G^{k(stb)} \implies F \not\equiv_{ND}^{\mathcal{E}_{stb}} G$.

1st case: Assume $A(F^{k(stb)}) \neq A(G^{k(stb)})$ and w.l.o.g. let $a \in A(F^{k(stb)}) \setminus A(G^{k(stb)})$. Since the stable kernel is node-preserving (Fact 5.12) we obtain $G \setminus B = (\emptyset, \emptyset)$ and $F \setminus B \in \{(\{a\}, \emptyset), (\{a\}, \{(a, a)\})\}$ if $B = (A(F) \cup A(G)) \setminus \{a\}$. In either case, $\emptyset \in \mathcal{E}_{stb}(G) \setminus \mathcal{E}_{stb}(F)$ since $\mathcal{E}_{stb}(F) \in \{\emptyset, \{\{a\}\}\}$. From now on we assume $A(F^{k(stb)}) = A(G^{k(stb)})$.

2nd case: Consider $R(F^{k(stb)}) \neq R(G^{k(stb)})$ and w.l.o.g. let $(a, b) \in R(F^{k(stb)}) \setminus R(G^{k(stb)})$. Let $a = b$. Remember that the stable kernel is loop-preserving (Fact 5.12). Therefore, $(a, a) \in R(F) \setminus R(G)$. We obtain $G \setminus B = (\{a\}, \emptyset)$ and $F \setminus B = (\{a\}, \{(a, a)\})$ if $B = (A(F) \cup A(G)) \setminus \{a\}$. Hence, $\emptyset = \mathcal{E}_{stb}(F) \neq \mathcal{E}_{stb}(G) = \{\{a\}\}$. From now on we assume $L(F^{k(stb)}) = L(G^{k(stb)})$. Consider now $a \neq b$. Consequently, $(a, b) \in R(F)$ and $(a, a) \notin R(F)$. Hence, $(a, a) \notin R(G)$ and furthermore, $(a, b) \notin R(G)$. Define $B = (A(F) \cup A(G)) \setminus \{a, b\}$. In any case, $\{a\} \in \mathcal{E}_{stb}(F \setminus B) \setminus \mathcal{E}_{stb}(G \setminus B)$ concluding the if-direction.

(\impliedby) Given $F^{k(stb)} = G^{k(stb)}$. Applying Theorems 5.14 and 5.20 one after the other yields $F \equiv_E^{\mathcal{E}_{stb}} G$ and then $F \equiv_N^{\mathcal{E}_{stb}} G$. Finally, Corollary 5.21 justifies $F \equiv_{ND}^{\mathcal{E}_{stb}} G$ concluding the proof. \square

¹³We mention that Theorem 10 in [Baumann, 2014a] hold, given that resulting AFs have to be non-empty. The claimed normal deletion equivalence of the AFs F and G depicted in [Baumann, 2014a, Example 4] can be disproved by setting $B = \{a, b, c, f\}$.

5.2.4 Characterization Theorems in Case of Self-loop-free AFs

We already observed that apart from naive kernel any mentioned kernel k does not change anything if the considered AF F is self-loop-free, i.e. $F = F^k$ (cf. Facts 5.12 and 5.18). Consequently, any equivalence relation characterizable through such a kernel collapses to identity if we restrict ourselves to self-loop-free AFs. This is stated in the following theorem.

Theorem 5.31. *Given any binary relation $\equiv \subseteq \mathcal{F} \times \mathcal{F}$ characterizable through k where $k \in \{k(stb), k(ad), k(gr), k(co), k^*(ad), k^*(gr), k^*(co), k^*(stg)\}$. For any self-loop-free AFs F and G ,*

$$F \equiv G \iff F = G.$$

We will refrain from listing all combinations of semantics and equivalence notions characterizable through a kernel mentioned in the theorem above. Please confer Figures 5.4 and 5.7 for compact overviews. For all such combinations, self-loop-free AFs are redundancy-free, i.e. all attacks as well as arguments may play a crucial role w.r.t. further evaluations and thus, there is no space for simplification. In the introductory part of this section we noted that many equivalence notions, e.g. normal and local expansion equivalence are motivated by the instantiation-based context where AFs are built from an underlying knowledge base. However, we want to mention that there are some formalisms like classical logic-based argumentation where self-attacking arguments do not occur [Besnard and Hunter, 2001, Theorem 4.13], while for other systems, e.g. ASPIC self-defeating arguments indeed may arise [Prakken, 2010, Section 7].

5.2.5 Summary of Results and Conclusion

In the presented results the notion of a kernel played a crucial role. Indeed, kernels are interesting from several perspectives: First, they allow to decide the corresponding notion of equivalence by a simple check for topological (i.e. syntactical) equality. Moreover, all kernels we have obtained so far can be efficiently constructed from a given argumentation framework. This means, if a certain equivalence notion is characterizable through such a kernel, then we have tractability of the associated decision problem.

The following Figure 5.4 provides a comprehensive overview of the state of the art in case of extension-based semantics. The entry “ k ” in row M and column σ indicates that \equiv_M^{σ} is characterizable through k . The abbreviation “ id ” stands for identity map and the question mark represents an open problem. Further abbreviations like “ L ” and “ Att^σ ” refer to additional conditions relevant in case of normal deletion equivalence (cf. Theorem 5.29). The entry “[m, n]” indicates three facts. First, the characterization problem is already solved in Theorem/Proposition n in m .¹⁴ Second, the characterization result is not (purely) kernel-based and third, it can be checked that none of the introduced kernels serve as a characterization.

Remember that any arbitrary expansion (deletion) can be split into a normal and local part. So one natural conjecture is that normal and local expansion (deletion) equivalence jointly imply expansion (deletion) equivalence. Using the results presented in this section we can not only verify the addressed conjecture but even give a significantly stronger result. In fact, the main and quite surprisingly relations for the considered semantics can be briefly and concisely stated in the following two equations, namely “normal expansion equivalence = expansion equivalence” and “local deletion equivalence = deletion equivalence”. The fact that different notions of equivalence might or might not coincide is interesting from

¹⁴For m we use the following assignments: 1 = [Baumann, 2011], 2 = [Baumann and Brewka, 2018], 3 = [Baumann and Brewka, 2013a] and 4 = [Oikarinen and Woltran, 2011]

	<i>stg</i>	<i>stb</i>	<i>ss</i>	<i>eg</i>	<i>ad</i>	<i>pr</i>	<i>il</i>	<i>gr</i>	<i>co</i>	<i>na</i>	<i>cf2</i>	<i>stg2</i>
W	?	[1,3]	?	?	[2,1]	[3,1]	?	?	[2,1]	?	?	?
L	$k^*(stg)$	[4,9]	$k(ad)$	$k(ad)$	$k(ad)$	$k(ad)$	$k(ad)$	[4,10]	[4,11]	$k(na)$?	?
E	$k(stb)$	$k(stb)$	$k(ad)$	$k(ad)$	$k(ad)$	$k(ad)$	$k(ad)$	$k(gr)$	$k(co)$	$k(na)$	<i>id</i>	<i>id</i>
N	$k(stb)$	$k(stb)$	$k(ad)$	$k(ad)$	$k(ad)$	$k(ad)$	$k(ad)$	$k(gr)$	$k(co)$	$k(na)$	<i>id</i>	<i>id</i>
S	$k(stb)$	$k(stb)$	$k(ad)$	$k(ad)$	$k^*(ad)$	$k^*(ad)$	$k^*(ad)$	$k^*(gr)$	$k^*(co)$	$k(na)$?	?
ND	?	$k(stb)$?	?	$k^*(ad)$ L, Att^{ad}	?	?	$k^*(gr)$ L, Att^{gr}	$k^*(co)$ L, Att^{co}	?	?	?
D	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>
LD	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>
U	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>

Figure 5.4: Extension-based Characterizations for Finite AFs

a conceptual point of view. To illustrate this let us have a look at normal and strong expansion equivalence. Recall that normal expansions add new arguments and possibly new attacks which involve at least one of the fresh arguments, while strong expansions (a subclass of normal expansions) restrict the possible attacks between the new arguments and the old ones to a single direction. In dynamic settings, both concepts can be justified in the sense that new arguments might be raised but this will not influence the relation between already existing arguments. For strong expansions, only strong arguments will be raised, i.e. arguments which cannot be attacked by existing ones. The corresponding equivalence notions now check whether two AFs are “equally robust” to such new arguments, and indeed, normal expansion equivalence always implies strong expansion equivalence but the other direction is only true for some of the semantics, namely stage, stable, semi-stable, eager and naive semantics. One interpretation is that when two AFs are not normal expansion equivalent, then this can be made explicit by only posing strong arguments (not attacked by existing ones), while for the other semantics this is not the case. For this particular example, it seems that the notion of admissibility which is more “explicit” in the admissible, preferred, ideal, grounded and complete semantics is responsible for the fact that frameworks might be strong expansion equivalent but not normal expansion equivalent.

In Figure 5.3 we presented preliminary relations between several notions of equivalence which hold for any semantics. The refinement depicted in Figure 5.5 applies to any extension-based semantics considered in this section.

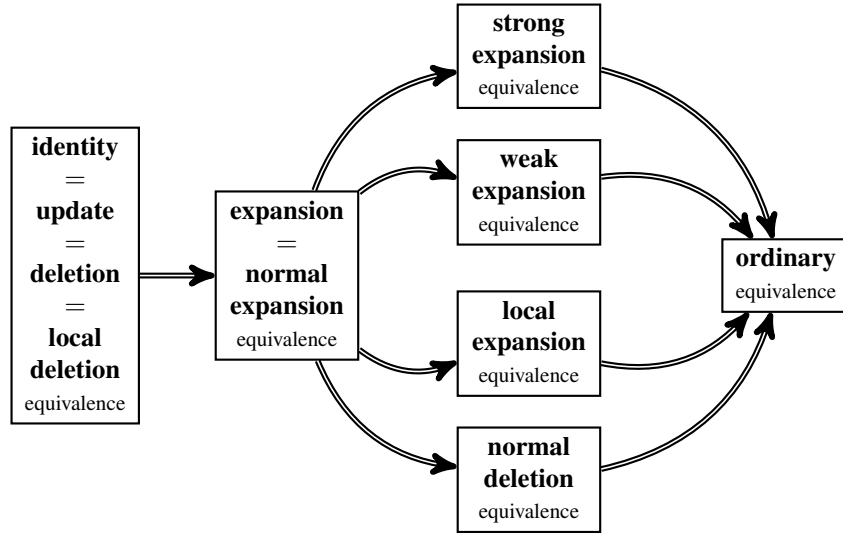


Figure 5.5: Relations for $\sigma \in \{stg, stb, ss, eg, ad, pr, il, gr, co, na, cf2, stg2\}$ - Extension-based Versions and Finite AFs

Finally, we present the overall picture for the most prominent semantics, namely the stable one. Interestingly, in contrast to Figure 5.5 all equivalence notions are comparable, i.e. they are totally ordered w.r.t. \subseteq . Comprehensive overviews for single semantics can be found in [Baumann, 2014b, Section 5.5.2] or [Baumann and Brewka, 2018]. The latter also contains a comparison to different notions of *minimal change equivalence* which are related to the so-called *minimal change problem* considered [Baumann, 2012b; Baumann and Brewka, 2013b]. As an aside, very recently the authors of [Baumann et al., 2017] introduced so-called *C-relativized equivalence* that subsumes ordinary and expansions equivalence as its extreme corner cases. The set C represents so-called *core* arguments which will not be directly touched by the possible expansions. This means, for any set C we obtain a further intermediate notion between expansion and ordinary equivalence. However, due to its recency further relations are not studied so far.

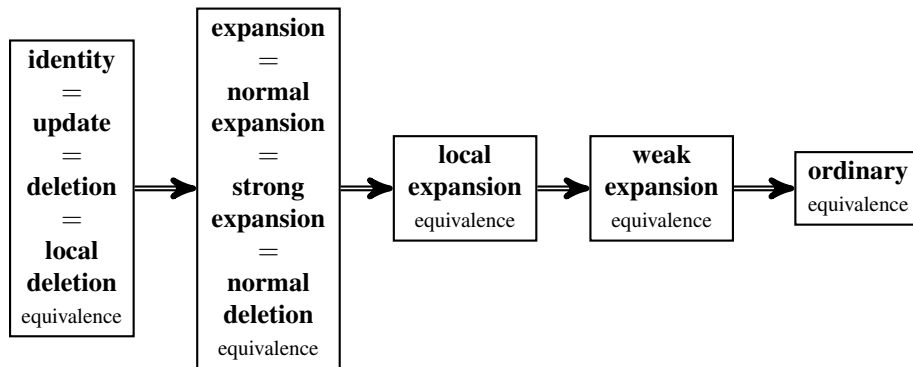


Figure 5.6: Stable Semantics - Extension-based Version and Finite AFs

5.3 Equivalence in the Light of Unrestricted Frameworks

Recently, a first study of several abstract properties in the unrestricted setting were presented in [Baumann and Spanring, 2017]. The main result regarding expansion equivalence can be summarized as follows: All characterization results carry over to the unrestricted setting as long as the AFs in question are *jointly expandable* (w.r.t. \mathcal{U}). Consider therefore the following definition and the corresponding characterization theorem.

Definition 5.32. F and G are jointly expandable if $\mathcal{U} \setminus (A(F) \cup A(G)) \neq \emptyset$.

Theorem 5.33. [Baumann and Spanring, 2017] For jointly expandable AFs F and G we have:

1. $F \equiv_E^{\mathcal{E}_\sigma} G \iff F^{k(\sigma)} = G^{k(\sigma)}$ for any $\sigma \in \{stb, ad, co, gr, na\}$,
2. $F \equiv_E^{\mathcal{E}_\sigma} G \iff F^{k(ad)} = G^{k(ad)}$ for any $\sigma \in \{pr, il, ss, eg\}$ and
3. $F \equiv_E^{\mathcal{E}_{stg}} G \iff F^{k(stb)} = G^{k(stb)}$.

The main proof strategies are straightforward extensions of those presented in [Oikarinen and Woltran, 2011]. However, finiteness assumptions are often used implicitly and one has to pay attention whether a certain reasoning step (e.g. subset relation between semantics, definedness statuses of semantics, finitely many extensions etc.) carry over to the infinite setting.

Interestingly, in case of the admissible as well as naive kernel we may even drop the restriction of joint expandability as stated in the following theorem.

Theorem 5.34. [Baumann and Spanring, 2017] For unrestricted AFs F and G we have:

1. $F \equiv_E^{\mathcal{E}_{na}} G \iff F^{k(na)} = G^{k(na)}$ and
2. $F \equiv_E^{\mathcal{E}_\sigma} G \iff F^{k(ad)} = G^{k(ad)}$ for any $\sigma \in \{ad, pr, il, ss, eg\}$.

The following two examples taken from [Baumann and Spanring, 2017] show that this assertion does not hold for all kernels considered in this section. The main reason for this different behaviour is that for some semantics it plays a decisive role whether AFs can be expanded by “fresh” arguments which is not given for unrestricted frameworks in general but guaranteed for jointly expandable AFs.

Example 5.35. Given $c \in \mathcal{U}$ and define the following two AFs, $F = (\mathcal{U} \setminus \{c\}, \{(a, a) \mid a \in \mathcal{U} \setminus \{c\}\})$ and $G = (\mathcal{U}, \{(a, a) \mid a \in \mathcal{U} \setminus \{c\}\})$. For any H we observe $\mathcal{E}_{stb}(F \dot{\cup} H) = \mathcal{E}_{stb}(G \dot{\cup} H)$. In particular,

$$\mathcal{E}_{stb}(F \dot{\cup} H) = \begin{cases} \{\{c\}\}, & \text{if } \{(c, a) \mid a \in \mathcal{U} \setminus \{c\}\} \subseteq R(H) \text{ and } (c, c) \notin R(H) \\ \emptyset, & \text{otherwise} \end{cases}$$

Consequently, $F \equiv_E^{\mathcal{E}_{stb}} G$ although $A(F) \neq A(G)$ (and thus, $F^{k(stb)} \neq G^{k(stb)}$).

Example 5.36. Consider the AFs $F = (\mathcal{U}, \{(a, a) \mid a \in \mathcal{U}\})$ and $G = (\mathcal{U}, \{(a, b) \mid a, b \in \mathcal{U}, a \neq b\})$. Applying the grounded kernel does not change anything for either framework, i.e. $F^{k(gr)} = F$ and $G = G^{k(gr)}$. Due to the absence of unattacked arguments we deduce $\mathcal{E}_{gr}(F \dot{\cup} H) = \mathcal{E}_{gr}(G \dot{\cup} H) = \{\emptyset\}$ for any AF H . Consequently, $F \equiv_E^{\mathcal{E}_{gr}} G$ although $F^{k(gr)} \neq G^{k(gr)}$.

5.4 Characterization Theorems for Labelling-Based Semantics

We now return to the finite setting and consider the second main approach used for evaluating argumentation scenarios, namely labelling-based semantics. As a matter of fact, the labelling-based versions of all considered semantics provides one with more information than their extension-based counterparts. More precisely, the defined 3-valued labellings assign a status to any argument of the considered AF F , i.e. in addition to the information which arguments are *accepted* we also have labels for the remaining arguments indicating that they are either *rejected* or *undecided* with respect to F . It is well known that many semantics establish a one-to-one correspondence between their extension-based and labelling-based versions (for more details see the subsequent Section 5.4.1). This means, any labelling is associated with exactly one extension and vice versa. It is not immediately apparent whether this property guarantees that there is a coincidence of the extension-based and labelling-based equivalence notions. In [Baumann, 2016] a negative answer was given. The main reason for the invalidity is that AFs may possess the same extensions without sharing the same arguments which is impossible in case of labellings since any argument has to be labelled. Furthermore, even sharing the same arguments does not ensure the validity of the converse direction. Consider therefore the following example.

Example 5.37. Consider the AFs F and G as depicted below. Although both frameworks possess the same unique preferred extension, they do not share the same preferred labellings. More precisely, $\mathcal{E}_{pr}(F) = \mathcal{E}_{pr}(G) = \{\{a\}\}$ but $\{(\{a\}, \{b\}, \emptyset)\} = \mathcal{L}_{pr}(F) \neq \mathcal{L}_{pr}(G) = \{(\{a\}, \emptyset, \{b\})\}$.



Moreover, observe that $F^{k^*(ad)} = G = G^{k^*(ad)}$. Consequently, both frameworks are even strong expansion equivalent w.r.t. preferred extension-based semantics (Theorem 5.20). This means, equivalence notions may differ considerably if considered under the extension-based or labelling-based approach.

In contrast to extension-based semantics where characterization results are spread over a high number of publications there is only one reference, namely [Baumann, 2016] concerned with labelling-based semantics. The author considered 8 different equivalence notions w.r.t. 8 prominent labelling-based semantics in the finite setting. In effect, similarly to extension-based semantics, almost all labelling-based equivalence notions can be decided syntactically. Differently from the extension-based approach we observe a much more homogeneous picture. For instance, there is no need for the more sophisticated σ -*-kernels as we will see.

5.4.1 Basic Properties and a Fundamental Relation

Before turning to the main results we start with some preliminary facts relating σ -extensions and σ -labellings. In the following we restrict ourselves to the semantics considered in [Baumann, 2016]. For any 3-valued labelling $\mathcal{L} = (\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ we use $\mathcal{L} = (\mathcal{L}^i, \mathcal{L}^o, \mathcal{L}^u)$ as usual.

Fact 5.38. Given a finite AF $F = (A, R)$ and $E \subseteq A$. We write $E^{\mathcal{L}}$ for $(E, E^+, A \setminus E^{\oplus})$. For all $\sigma \in \{stb, ss, eg, ad, pr, il, gr, co\}$ we have,

1. If $\mathcal{L} \in \mathcal{L}_{\sigma}(F)$, then $\mathcal{L}^i \in \mathcal{E}_{\sigma}(F)$, (extension induced by labelling)
2. If $E \in \mathcal{E}_{\sigma}(F)$, then $E^{\mathcal{L}} \in \mathcal{L}_{\sigma}(F)$ and (labelling induced by extension)
3. Obviously, $(E^{\mathcal{L}})^i = E$. ($I \circ \mathcal{L} = id$)

We point out that the first two properties mentioned in Fact 5.38 do not ensure that there is a one-to-one correspondence between σ -labellings and σ -extensions. This desirable feature (which would

indeed justify the terms σ -labellings and σ -extensions) is given if additionally, labellings are uniquely determined by their in-labelled arguments.

Fact 5.39. *Given a finite AF $F = (A, R)$ and a set $E \subseteq A$. For all semantics $\sigma \in \{stb, ss, eg, pr, il, gr, co\}$ we have,*

1. *For any $\mathcal{L}, M \in \mathcal{L}_\sigma(F)$, $\mathcal{L}' = M'$ iff $\mathcal{L} = M$, (uniquely determined by in-labels)*
2. *Given $\mathcal{L} \in \mathcal{L}_\sigma(F)$, then $(\mathcal{L}')^\mathcal{L} = \mathcal{L}$ and ($\mathcal{L} \circ I = id$)*
3. *$|\mathcal{L}_\sigma(F)| = |\mathcal{E}_\sigma(F)|$. (same cardinality)*

As an aside, we mention that (although not immediately apparent) the first two items of Fact 5.39 are equivalent independently of any semantics definition. Please note that admissible labellings are excluded from Fact 5.39. The AF F depicted in Example 5.37 shows that this is no coincidence. It possesses two admissible labellings associated with one admissible extension. More precisely, the admissible labellings $(\{a\}, \{b\}, \emptyset)$ as well as $(\{a\}, \emptyset, \{b\})$ refer to the same admissible extension $\{a\}$.

We proceed with a general relation between labelling-based and extension-based versions of certain equivalence notion. More precisely, for any considered semantics and any equivalence notion presented in Definition 5.6 we have that being equivalent w.r.t. labellings implies being equivalent w.r.t. extensions. The main reason for this fundamental relation is the following lemma stating that possessing the same labellings implies sharing the same extensions. We mention that this property is already guaranteed if the semantics σ in question satisfies that any σ -extension induces an σ -labelling and vice versa (cf. statements 1 and 2 of Fact 5.38).

Lemma 5.40 ([Baumann, 2016]). *Given two finite AFs F, G . For any $\sigma \in \{stb, ss, eg, ad, pr, il, gr, co\}$ we have,*

$$\mathcal{L}_\sigma(F) = \mathcal{L}_\sigma(G) \implies \mathcal{E}_\sigma(F) = \mathcal{E}_\sigma(G).$$

Proof. Reductio ad absurdum. Assume $\mathcal{E}_\sigma(F) \neq \mathcal{E}_\sigma(G)$. Then, w.l.o.g. exists $E \in \mathcal{E}_\sigma(F) \setminus \mathcal{E}_\sigma(G)$. Consequently, $E^\mathcal{L} \in \mathcal{L}_\sigma(F)$ (item 2 of Fact 5.38). Thus, $E^\mathcal{L} \in \mathcal{L}_\sigma(G)$ (assumption). Hence, $(E^\mathcal{L})^I \in \mathcal{E}_\sigma(G)$ (Item 1 of Fact 5.38). Furthermore, $(E^\mathcal{L})^I = E \in \mathcal{E}_\sigma(G)$ (item 3 of Fact 5.38). Contradiction! \square

We now present the fundamental relation between labelling-based and extension-based equivalence notion.

Theorem 5.41 ([Baumann, 2016]). *Given two finite AFs F and G . For any $\sigma \in \{stb, ss, eg, ad, pr, il, gr, co\}$ and any $M \in \{W, L, E, N, S, ND, D, LD, U\}$ we have,*

$$F \equiv_M^{\mathcal{L}_\sigma} G \implies F \equiv_M^{\mathcal{E}_\sigma} G.$$

Proof. We show the contrapositive. Assume $F \not\equiv_M^{\mathcal{E}_\sigma} G$. This means, there is a certain scenario S according to M , s.t. $\mathcal{E}_\sigma(S(F)) \neq \mathcal{E}_\sigma(S(G))$.¹⁵ Consequently, $\mathcal{L}_\sigma(S(F)) \neq \mathcal{L}_\sigma(S(G))$ (Lemma 5.40) proving $F \not\equiv_M^{\mathcal{L}_\sigma} G$. \square

In Example 5.37 we have seen that the converse direction does not hold in general. Nevertheless, there is huge number of equivalence notions where labelling-based and extension-based versions do indeed coincide (cf. Figure 5.7 for an overview).

¹⁵For instance, in case of expansion equivalence (i.e. $M = E$) a scenario S is simply the union with a further AF H , i.e. $S(F) = F \dot{\cup} H$ and $S(G) = G \dot{\cup} H$.

5.4.2 Coincidences of Extension-based and Labelling-based Versions

Remember that the identity relation is the finest equivalence relation. Furthermore, it is already shown that deletion, local deletion as well as update equivalence w.r.t. \mathcal{E}_σ collapse to identity (see Figure 5.5). Consequently, applying the fundamental relation stated in Theorem 5.41 we obtain the identical characterization results w.r.t. labelling-based semantics.

Theorem 5.42 ([Baumann, 2016]). *For finite AFs F and G , a scenario $M \in \{D, LD, U\}$ and a semantics $\sigma \in \{stb, ss, eg, ad, pr, il, gr, co\}$ we have,*

$$F \equiv_M^{\mathcal{L}_\sigma} G \iff F = G.$$

Analogously to extension-based semantics (cf. Fact 5.13) we have that there are combinations of kernels and semantics σ , s.t. the application of a kernel does not vary the set of σ -labellings.

Fact 5.43. *For any finite AF F ,*

1. $\mathcal{L}_\sigma(F) = \mathcal{L}_\sigma(F^{k(\sigma)})$ for $\sigma \in \{co, stb, gr\}$ and
2. $\mathcal{L}_\tau(F) = \mathcal{L}_\tau(F^{k(ad)})$ for $\tau \in \{ss, eg, pr, il\}$.

The fact above is the decisive property which allows one to carry over further kernel-based characterization results for extension-based semantics to their labelling-based version. In order to show this result it was necessary to find a condition for equality of two complete labellings of different AFs. Remember that two complete labellings of the same framework are identical if and only if they possess the same in-labelled arguments (Fact 5.39). In case of different AFs we have to require additionally that both frameworks share the same arguments and the same range w.r.t. the set of in-labelled arguments.

Fact 5.44. *Given two finite AFs F and G as well as $\mathcal{L} \in \mathcal{L}_{co}(F)$ and $M \in \mathcal{L}_{co}(G)$. We have $\mathcal{L} = M$ iff simultaneously $A(F) = A(G)$, $\mathcal{L}' = M'$ and $R_F^+(\mathcal{L}') = R_G^+(M')$.*

Please observe that admissible labellings do not fulfill Fact 5.44. Consider for instance again the AF F depicted in Example 5.37 and its two admissible labellings $(\{a\}, \{b\}, \emptyset)$ and $(\{a\}, \emptyset, \{b\})$.

We proceed with the main coincidence theorem. It stipulates that several expansion equivalence relations as well as weaker notions do not distinguish between their labelling-based and extension-based version. This means, kernel-based characterization results (depicted in Figure 5.3) carry over to labelling-based semantics. Similarly to extension-based semantics we present an overview of characterizing kernels at the end of this section (cf. Figure 5.7).

Theorem 5.45 ([Baumann, 2016]). *Given finite AFs F and G . We have,*

1. $F \equiv_M^{\mathcal{E}_\sigma} G \iff F \equiv_M^{\mathcal{L}_\sigma} G$ for $\sigma \in \{stb, ss, eg, pr, il, gr, co\}, M \in \{E, N\}$,
2. $F \equiv_L^{\mathcal{E}_\sigma} G \iff F \equiv_L^{\mathcal{L}_\sigma} G$ for $\sigma \in \{ss, eg, pr, il\}$ and
3. $F \equiv_S^{\mathcal{E}_\sigma} G \iff F \equiv_S^{\mathcal{L}_\sigma} G$ for $\sigma \in \{stb, ss, eg\}$.

5.4.3 Non-Coincidence of Extension-based and Labelling-based Versions

We now leave the realm of uniformity of extension-based and labelling-based characterizations. This section is divided into three parts. We start with characterization theorems for admissible labellings. In particular, we will see that the admissible kernel (originally introduced to characterize equivalence

notions w.r.t. admissible extension-based semantics) does not serve as characterizing kernel for admissible labellings. We then proceed with strong expansion equivalence w.r.t. labellings. We will see that the remaining notions are characterizable via traditional kernels instead of σ -*-kernels. In the third part we consider normal deletion equivalence w.r.t. labelling-based semantics. In contrast to their extension-based versions where many notions has defied any attempt of solving, we present characterization theorems based on traditional kernels for all eight considered semantics.

5.4.3.1 Expansion Equivalence w.r.t. Admissible Labellings

Expansion equivalence as well as its local, normal and strong versions w.r.t. admissible extensions are characterizable through the admissible kernel. The following example shows that this assertion does not hold in case of admissible labellings.

Example 5.46. The following two AFs possess the same admissible kernels, namely $F^{k(ad)} = G^{k(ad)} = F$. Consequently, applying characterization theorems for extension-based semantics we obtain $F \equiv_M^{\mathcal{L}_{ad}} G$ for $M \in \{L, E, N\}$ (cf. Figure 5.4).



Observe that $(\{b\}, \emptyset, \{a\}) \in \mathcal{L}_{ad}(G) \setminus \mathcal{L}_{ad}(F)$ because the argument a cannot be undecided in F since it attacks the in-labelled argument b . Thus $F \not\equiv_M^{\mathcal{L}_{ad}} G$ for $M \in \{L, E, N, S\}$.

Let us assume that the equivalence notions considered in the example above are characterizable through a certain kernel k . Due to the fundamental relation (Theorem 5.41) and the characterization results w.r.t. admissible extensions (Figure 5.4), we already know that the kernel k has to satisfy the following implication: $F^k = G^k \implies F^{k(ad)} = G^{k(ad)}$ for any two AFs F and G . This means, we are looking for a weaker kernel than the admissible one in the sense that first, everything which is redundant w.r.t. k has to be redundant w.r.t. the admissible kernel too; and second, an attack from a to b has to survive even if a is self-defeating and b counterattacks a . One candidate for k is the complete kernel since redundancy w.r.t. the complete kernel implies redundancy w.r.t. to the admissible one, and furthermore, it deletes an attack between two arguments if and only if both are self-defeating. And indeed, it was shown that expansion equivalence as well as its local, normal and strong variant w.r.t. admissible labellings are characterizable through the complete kernel as stated by the following theorem.

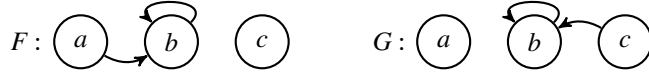
Theorem 5.47 ([Baumann, 2016]). *Given finite AFs F and G . We have,*

$$F \equiv_M^{\mathcal{L}_{ad}} G \iff F^{k(co)} = G^{k(co)} \text{ with } M \in \{L, E, N, S\}.$$

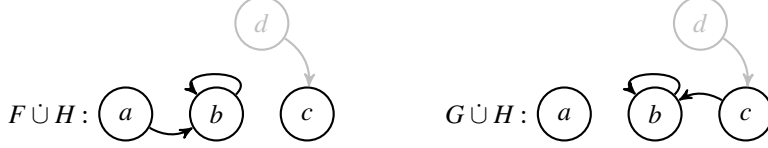
5.4.3.2 Strong Expansion Equivalence for Preferred, Ideal, Grounded and Complete Labellings

In this subsection we present characterization theorems for strong expansion equivalence w.r.t. labelling-based preferred, ideal, grounded and complete semantics. Remember that in case of strong expansions a former attack between old arguments will never become a counterattack to an added attack. Consequently, in contrast to arbitrary expansions former attacks do not play a role with respect to being a potential defender of an added argument. The context-sensitive σ -*-kernels took these considerations into account and allow for more deletions than their classical counterparts.

Example 5.48. According to Definition 5.17 we have, $F^{k^*(\sigma)} = G^{k^*(\sigma)}$ for any semantics $\sigma \in \{ad, gr, co\}$. More precisely, the attacks (a, b) in F as well as (c, b) in G are redundant w.r.t. all three σ -*-kernels. This means, in consideration of Figure 5.4 both frameworks are strong expansion equivalent w.r.t. the extension-based versions of preferred, ideal, grounded and complete semantics.



Consider the following dynamic scenario where a stronger argument than the former ones is added. Formally, we conjoin the AF $H = (\{c, d\}, \{(d, c)\})$ to both frameworks F and G .



Note that both frameworks has to possess the same σ -extension since $G \equiv_S^{\mathcal{L}_\sigma} H$ for $\sigma \in \{pr, il, gr, co\}$ is already ensured. Furthermore, we observe $(\{a, d\}, \{b, c\}, \emptyset) \in \mathcal{L}_\sigma(F \cup H) \setminus \mathcal{L}_\sigma(G \cup H)$ since b cannot be out-labelled in $G \cup H$ because there is no in-labelled attacker. This means, $F \not\equiv_S^{\mathcal{L}_\sigma} G$ for $\sigma \in \{pr, il, gr, co\}$.

Analogously to the previous section let us assume that strong expansion equivalence w.r.t. the considered labelling-based semantics are characterizable through a certain kernel k . We immediately obtain, $F^k = G^k \implies F^{k^*(\sigma)} = G^{k^*(\sigma)}$ for any two AFs F and G . Possible candidates are the classical counterparts of the σ -*-kernels and indeed it was shown that these kernels guarantee the desired outcome. This means, in case of strong expansion equivalence w.r.t. preferred, ideal, grounded and complete semantics we have that the labelling-based version is characterizable through a classical σ -kernel if and only if the extension-based version is characterizable through the corresponding σ -*-kernel.

Theorem 5.49 ([Baumann, 2016]). *Given finite AFs F and G . We have,*

1. $F \equiv_S^{\mathcal{L}_\sigma} G \iff F^{k(ad)} = G^{k(ad)}$ for $\sigma \in \{pr, il\}$,
2. $F \equiv_S^{\mathcal{L}_{gr}} G \iff F^{k(gr)} = G^{k(gr)}$ and
3. $F \equiv_S^{\mathcal{L}_{co}} G \iff F^{k(co)} = G^{k(co)}$.

5.4.3.3 Normal Deletion Equivalence

Characterizing normal deletion equivalence in case of extension-based semantics is exceptional in several regards. Remember that normal deletions retract arguments and their corresponding attacks. Firstly, only a few characterization results are achieved (cf. Figure 5.4). Furthermore, apart from stable semantics, none of the characterization results is purely kernel-based, i.e. beside the equality of kernels on certain parts of the frameworks further loop- as well as attack-conditions have to be satisfied. Finally, quite surprisingly, normal deletion equivalent AFs do not even have to share the same arguments enabling equivalence classes with an infinite number of elements. Being equivalent w.r.t. labellings and possessing different arguments at the same time is impossible in case of labellings since any argument has to be labelled. It turned out that any considered labelling-based semantics is characterizable through traditional kernels and thus, do not share any of the features mentioned above. Consider the following main theorem.

Theorem 5.50 ([Baumann, 2016]). *Given finite AFs F and G . We have,*

1. $F \equiv_{ND}^{\mathcal{L}_{stb}} G \iff F^{k(stb)} = G^{k(stb)}$,
2. $F \equiv_{ND}^{\mathcal{L}_\sigma} G \iff F^{k(ad)} = G^{k(ad)}$ for $\sigma \in \{ss, eg, pr, il\}$,
3. $F \equiv_{ND}^{\mathcal{L}_\sigma} G \iff F^{k(co)} = G^{k(co)}$ for $\sigma \in \{ad, co\}$ and
4. $F \equiv_{ND}^{\mathcal{L}_{gr}} G \iff F^{k(gr)} = G^{k(gr)}$.

5.4.4 Summary of Results and Conclusion

The following Figure 5.7 presents a comprehensive overview of the state of the art in case of labelling-based semantics. Analogously to Figure 5.4 the entry “ k ” in row M and column σ indicates that $\equiv_M^{\mathcal{L}\sigma}$ is characterizable through k given the finiteness restriction. The abbreviation “ id ” stands for identity map and the question mark represents an open problem.¹⁶ A grey-highlighted entry reflects the situation that extension-based and labelling-based version do not coincide.

	<i>stb</i>	<i>ss</i>	<i>eg</i>	<i>ad</i>	<i>pr</i>	<i>il</i>	<i>gr</i>	<i>co</i>
L	?	$k(ad)$	$k(ad)$	$k(co)$	$k(ad)$	$k(ad)$?	?
E	$k(stb)$	$k(ad)$	$k(ad)$	$k(co)$	$k(ad)$	$k(ad)$	$k(gr)$	$k(co)$
N	$k(stb)$	$k(ad)$	$k(ad)$	$k(co)$	$k(ad)$	$k(ad)$	$k(gr)$	$k(co)$
S	$k(stb)$	$k(ad)$	$k(ad)$	$k(co)$	$k(ad)$	$k(ad)$	$k(gr)$	$k(co)$
ND	$k(stb)$	$k(ad)$	$k(ad)$	$k(co)$	$k(ad)$	$k(ad)$	$k(gr)$	$k(co)$
D	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>
LD	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>
U	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>	<i>id</i>

Figure 5.7: Labelling-based Characterizations for Finite AFs

In contrast to extension-based semantics we observe a much more homogeneous picture. Firstly, there is no need for the more sophisticated σ -*-kernels. Secondly, normal deletion equivalence w.r.t. labelling-based semantics is naturally incorporated in the overall picture in the sense that it coincides with its corresponding expansion, normal expansion and strong expansion equivalence notions.

The following Figure 5.8 applies to each one of the eight labelling-based semantics considered in this section. In comparison to Figure 5.3 where preliminary relations are depicted it illustrates (to a certain extent) a collapse of the diversity of the introduced equivalence notions in case of labelling-based semantics.

¹⁶In contrast to extension-based semantics the labelling versions of conflict-free-based semantics like stage, naive, cf2 as well as stage2 semantics (cf. [Caminada, 2011; Gaggl and Dvořák, 2016]) as well as weak expansion equivalence at all were not considered so far and thus, represent open problems too.

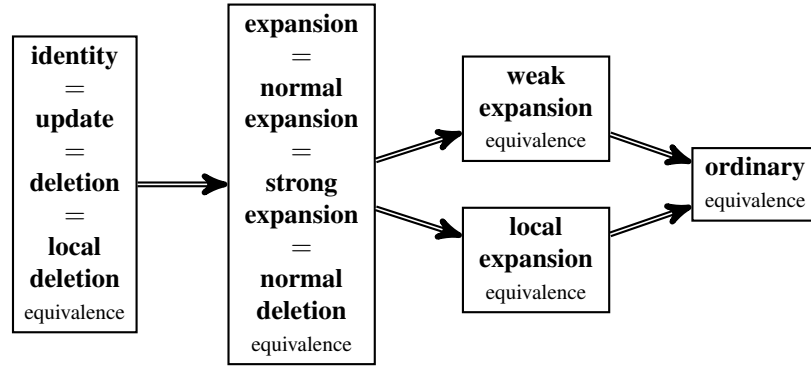


Figure 5.8: Relations for $\sigma \in \{stb, ss, eg, ad, pr, il, gr, co\}$ - Labelling-based Versions and Finite AFs

5.5 Final Remarks

In this chapter we motivated and discussed several notions of equivalence in the context of abstract argumentation and provided an exhaustive number of characterization theorems for extension-based as well as labelling-based semantics. In general we may state that Dung's abstract argumentation frameworks are a very compact formalism since the majority of the considered equivalence notions possess only little space for redundancy. Moreover, most of these notions collapse to identity if self-loop-free AFs are considered. This means, in this case any subframework of the AF in question may play a decisive role w.r.t. further evaluations and thus, cannot be locally replaced by another. This insight is sometimes used as an argument against the usefulness of the study of equivalence notions in the context of abstract argumentation. Obviously, we agree that if you are expecting much space for simplification, then the results are somehow disappointing but let us not lose sight of the fact that this is only clear *after* it has been proved. Furthermore, as already stated, the results underline that in case of abstract argumentation (almost) everything is meaningful similar to other non-monotonic formalisms available in the literature (cf. [Lifschitz et al., 2001] for logic programs, [Turner, 2004] for causal theories, [Turner, 2001] for default logic and [Truszczyński, 2006] for nonmonotonic logics in general). However, one decisive difference to these formalisms is that equivalence notions in case of abstract argumentation can be decided syntactically. Indeed, kernels are interesting from several perspectives: First, they allow to decide the corresponding notion of equivalence by a simple check for topological equality and second, all kernels we have obtained so far can be efficiently constructed from a given argumentation framework. This means, if a certain equivalence notion is characterizable through such a kernel, then we have tractability of the associated decision problem.

Chapter 6

Verifiability of Argumentation Semantics

Over the last 20 years a series of abstract argumentation semantics were introduced. The motivations of these semantics range from the desired treatment of specific examples to fulfilling a number of abstract principles. The comparison via abstract criteria of the different semantics available is a topic which emerged quite recently in the community ([Baroni and Giacomin, 2007b] can be seen as the first paper in this line). In this chapter we take a further step towards a comprehensive understanding of argumentation semantics. In particular, we study the following question: Do we really need the entire AF F to compute a certain argumentation semantics σ ? In other words, is it possible to unambiguously determine acceptable sets w.r.t. σ , given only partial information of the underlying framework F . In order to solve this problem let us start with the following reflections:

1. As a matter of fact, one basic requirement of almost all existing semantics (exemptions are given in [Jakobovits and Vermeir, 1999; Arieli, 2012; Grossi and Modgil, 2015]) is that of conflict-freeness, i.e. arguments within a reasonable position are not allowed to attack each other. Consequently, knowledge about conflict-free sets is an essential part for computing semantics.
2. The second step is to ask the following: Which information on top on conflict-free sets has to be added? Imagine the set of conflict-free sets given by $\{\emptyset, \{a\}, \{b\}\}$. Consequently, there has to be at least one attack between a and b . Unfortunately, this information is not sufficient to compute any standard semantics (except naive extensions, which are defined as \subseteq -maximal conflict-free sets) since we know nothing precise about the neighborhood of a and b . The following three AFs possess exactly the mentioned conflict-free sets, but differ with respect to other semantics.



3. The final step is to try to minimize the added information. In other words, which kind of knowledge about the neighborhood is somehow dispensable in the light of computation? Clearly, this will depend on the considered semantics. For instance, in case of stage semantics [Verheij, 1996], which requests conflict-free sets of maximal range, we do not need any information about incoming attacks. This information can not be omitted in case of admissible-based semantics since incoming attacks require counterattacks.

The above considerations motivate the introduction of *verification classes* specifying a certain amount of information. In a first step, we study the relation of these classes to each other. We therefore introduce the

notion of being *more informative*, capturing the intuition that a certain class can reproduce the information of another. We present a hierarchy w.r.t. this ordering, containing 15 different verification classes only. This is because many syntactically different classes collapse to the same amount of information.

We then formally define the essential property of a semantics σ being *verifiable* w.r.t. a certain verification class. We present a general theorem stating that any *rational* semantics is exactly verifiable w.r.t. one of the 15 different verification classes. Roughly speaking, a semantics is rational if attacks inbetween two self-loops can be omitted without affecting the set of extensions. An important aside hereby is that even the most informative class contains indeed less information than the entire framework by itself.

We consider a representative set of standard semantics. All of them satisfy rationality and thus, are exactly verifiable w.r.t. a certain class. Since the theorem does not provide an answer to which verification class perfectly matches a certain rational semantics we study this problem one by one for any considered semantics. As a result, only six different classes are essential to classify the considered standard semantics.

In the last section we study an application of the concept of verifiability. More precisely, we address the question of strong equivalence for semantics lying inbetween known semantics, so-called *intermediate semantics*. Strong equivalence is the natural counterpart to ordinary equivalence in monotonic theories (see [Oikarinen and Woltran, 2011; Baumann, 2016] for abstract argumentation and [Maher, 1986; Lifschitz et al., 2001; Turner, 2004; Truszczyński, 2006] for other nonmonotonic theories). We provide characterization theorems relying on the notion of verifiability and thus, contributing to a more abstract understanding of the different features argumentation semantics offer.

6.1 Preliminaries and Strongly Admissible Sets

In this chapter we will consider finite AFs and extension-based semantics only. In Section 3.1.2 we introduced the so-called *range* E_F^\oplus of a certain set E w.r.t. a given AF $F = (A, R)$. It is given by $E_F^\oplus = E \cup E_F^+$ where $E_F^+ = \{b \mid (a, b) \in R, a \in E\}$. Similarly, we will use E_F^\ominus to denote the *anti-range* of E defined as $E \cup E_F^-$ with $E_F^- = \{b \mid (b, a) \in R, a \in E\}$. If clear from context we drop the indices and simply write E^\oplus , E^+ , E^\ominus or E^- respectively.

We now introduce the concepts of *intermediate semantics* and *rationality* which will play a decisive role in this chapter. Remember that we use $\sigma \subseteq \tau$ to indicate that $\sigma(F) \subseteq \tau(F)$ for each AF F .

Definition 6.1. Given three semantics ρ, σ, τ . If we have $\rho \subseteq \sigma$ and $\sigma \subseteq \tau$, we say that σ is ρ - τ -intermediate.

Several examples of intermediate semantics can be found in Proposition 3.6. For instance, we have that *ss* is *stb-pr*-intermediate.

Definition 6.2. A semantics σ *rational* if for each AF F , $\sigma(F) = \sigma(F^l)$. We have $F^l = (A(F), R(F) \setminus \{(a, b) \in R(F) \mid (a, a), (b, b) \in R(F), a \neq b\})$.

Indeed, all semantics introduced in Definition 3.4 are rational. A prominent semantics that is based on conflict-free sets, but is not rational is the *cf2*-semantics [Baroni et al., 2005], since here chains of self-loops can have an influence on the SCCs of an AF (see also [Gaggl and Woltran, 2013]).

In order to present an exhaustive analysis of intermediate semantics (cf. Section 6.3) we provide a missing characterizing kernel for strongly admissible sets firstly introduced in [Baroni and Giacomin, 2007b]. We will see that, besides grounded [Oikarinen and Woltran, 2011] and resolution-based grounded semantics [Baroni et al., 2011b; Dvorák et al., 2014], strongly admissible sets are characterizable through the grounded kernel. Consider the following self-referential definition taken from [Caminada, 2014].

Definition 6.3. Given an AF $F = (A, R)$. A set $S \subseteq A$ is *strongly admissible*, i.e. $S \in \text{sad}(F)$ iff any $a \in S$ is defended by a strongly admissible set $S' \subseteq S \setminus \{a\}$.

The following properties are needed to prove the characterization theorem. (1) and (2) are already shown in [Baroni and Giacomin, 2007b], (3) is an immediate consequence of the former.

Proposition 6.4. *Given two AFs F and G , it holds that*

1. $gr(F) \subseteq sad(F) \subseteq ad(F)$,
2. if $S \in gr(F)$ we have: $S' \subseteq S$ for all $S' \in sad(F)$, and
3. $sad(F) = sad(G)$ implies $gr(F) = gr(G)$.

We now provide an alternative criterion for being a strongly admissible set. In contrast to the former it allows one to construct strongly admissible sets step by step which enables a construction method.

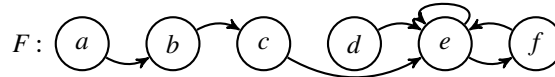
Definition 6.5. Given an AF $F = (A, R)$. A set $S \subseteq A$ is *strongly admissible*, i.e. $S \in sad(F)$ iff there are finitely many and pairwise disjoint sets A_1, \dots, A_n , s.t. $S = \bigcup_{1 \leq i \leq n} A_i$ and $A_1 \subseteq \Gamma_F(\emptyset)^1$ and furthermore, $\bigcup_{1 \leq i \leq j} A_i$ defends A_{j+1} for $1 \leq j \leq n-1$.

The following proof shows that both definitions are indeed equivalent.

Proposition 6.6. *Definitions 6.3 and 6.5 are equivalent.*

Proof. For the proof we use $S \in sad_k(F)$ as a shorthand for $S \in sad(F)$ in the sense of Definition k .
 (\Leftarrow) Given $S \in sad_{6.5}(F)$. Hence, there is a finite partition, s.t. $S = \bigcup_{1 \leq i \leq n} A_i$, $A_1 \subseteq \Gamma_F(\emptyset)$ and $\bigcup_{1 \leq i \leq j} A_i$ defends A_{j+1} for $1 \leq j \leq n-1$. Observe that $\bigcup_{1 \leq i \leq j} A_i \in sad_{6.5}(F)$ for any $j \leq n$. Let $a \in S$. Consequently, there is an index i^* , s.t. $a \in A_{i^*}$. Furthermore, since $\bigcup_{1 \leq i \leq i^*-1} A_i$ defends A_{i^*} by definition, we deduce that $\bigcup_{1 \leq i \leq i^*-1} A_i \subseteq S \setminus \{a\}$ defends a . We have to show now that (the smaller set w.r.t. \subseteq) $\bigcup_{1 \leq i \leq i^*-1} A_i \in sad_{6.3}(F)$. Note that $\bigcup_{1 \leq i \leq i^*-1} A_i \in sad_{6.5}(F)$. Since we are dealing with finite AFs we may iterate our construction. Hence, no matter which elements are chosen we end up with a \subseteq -chain, s.t. $\emptyset \subseteq \bigcup_{1 \leq i \leq i_e} A_i \subseteq S_e \setminus a_e$ and \emptyset defends a_e for some index i_e , set S_e and element a_e . This means, the question whether $S \in sad_{6.3}(F)$ can be decided positively by proving $\emptyset \in sad_{6.3}(F)$. Since the empty set does not contain any elements we find $\emptyset \in sad_{6.3}(F)$ concluding $sad_{6.5} \subseteq sad_{6.3}$.
 (\Rightarrow) Given $S \in sad_{6.3}(F)$, consider the following sets S_i : $S_1 = (\Gamma(\emptyset) \setminus \emptyset) \cap S$, $S_2 = (\Gamma(S_1) \setminus S_1) \cap S$, $S_3 = (\Gamma(\bigcup_{i=1}^2 S_i) \setminus \bigcup_{i=1}^2 S_i) \cap S$, \dots , $S_n = (\Gamma(\bigcup_{i=1}^{n-1} S_i) \setminus \bigcup_{i=1}^{n-1} S_i) \cap S$. Since we are dealing with finite AFs there has to be a natural $n \in \mathbb{N}$, s.t. $S_n = S_{n+1} = S_{n+2} = \dots$. Consider now the union of these sets, i.e. $\bigcup_{i=1}^n S_i$. We show now that $\bigcup_{i=1}^n S_i \in sad_{6.5}(F)$ and $\bigcup_{i=1}^n S_i = S$. By construction we have $S_1 \subseteq \Gamma(\emptyset)$. Moreover, $\bigcup_{1 \leq i \leq j} S_i$ defends S_{j+1} for $1 \leq j \leq n-1$. This can be seen as follows. By definition $S_{j+1} = (\Gamma(\bigcup_{i=1}^j S_i) \setminus \bigcup_{i=1}^j S_i) \cap S$. This means, $S_{j+1} \subseteq \Gamma(\bigcup_{i=1}^j S_i)$. Since $\Gamma(\bigcup_{i=1}^j S_i)$ contains all elements defended by $\bigcup_{i=1}^j S_i$ we obtain $\bigcup_{i=1}^n S_i \in sad_{6.5}(F)$. Obviously, $\bigcup_{i=1}^n S_i \subseteq S$. In order to derive a contradiction we suppose $S \not\subseteq \bigcup_{i=1}^n S_i$. This means there is a nonempty set S^* , s.t. $S = S^* \cup \bigcup_{i=1}^n S_i$. Let $S^* = \{s_1, \dots, s_k\}$. Observe that no element s_i is defended by $\bigcup_{i=1}^n S_i$ (*). Since $S \in sad_{6.3}(F)$ we obtain a set $S_1^* \subseteq S \setminus \{s_1\}$, s.t. $S_1^* \in sad_{6.3}(F)$ and S_1^* defends s_1 . We now iterate this procedure ending up with a set $S_k^* \subseteq S_{k-1}^* \setminus \{s_k\} \subseteq \bigcup_{i=1}^n S_i$, s.t. $S_k^* \in sad_{6.3}(F)$ and S_k^* defends s_k contradicting (*) and concluding the proof. \square

Example 6.7. Consider the following AF F .



¹Hereby, Γ is the so-called *characteristic function* (cf. Section 3.1.2). Note that the term $\Gamma_F(\emptyset)$ can be equivalently replaced by $\{a \in A \mid a \text{ is unattacked}\}$.

We have $\Gamma_F(\emptyset) = \{a, d\}$. Hence, for all $S \subseteq \{a, d\}$, $S \in \text{sad}(F)$. Furthermore, $\Gamma_F(\{a\}) = \{a, c\}$, $\Gamma_F(\{d\}) = \{d, f\}$ and $\Gamma_F(\{a, d\}) = \{a, d, c, f\}$. This means, additionally $\{a, c\}, \{d, f\}, \{a, d, c\}, \{a, d, f\}, \{a, d, c, f\} \in \text{sad}(F)$. Finally, $\Gamma_F(\{a, c\}) = \{a, c, f\}$ justifying the last missing set $\{a, c, f\} \in \text{sad}(F)$.

The following corollary is an immediate consequence of Definition 6.5. It is essential to prove the characterization theorem for strongly admissible sets.

Corollary 6.8. *Given an AF F and two sets $B, B' \subseteq A(F)$. If B defends B' , then $B \cup B'$ is strongly admissible if B is.*

The grounded kernel is insensitive w.r.t. strongly admissible sets, which then allows us to state the main result for strongly admissible sets.

Lemma 6.9. *For any AF F , $\text{sad}(F) = \text{sad}(F^{k(\text{gr})})$.*

Proof. The grounded kernel is node- and loop-preserving, i.e. $A(F) = A(F^{k(\text{gr})})$ and $L(F) = L(F^{k(\text{gr})})$. Furthermore, $\text{cf}(F) = \text{cf}(F^{k(\text{gr})})$ and $\Gamma_F(\emptyset) = \Gamma_{F^{k(\text{gr})}}(\emptyset)$ as shown in [Oikarinen and Woltran, 2011, Lemma 6].

(\subseteq) Given $S \in \text{sad}(F)$. The proof is by induction on n indicating the number of sets forming a suitable (according to Definition 6.5) partition of S . Let $n = 1$. In consideration of the grounded kernel we observe $\Gamma_F(\emptyset) = \Gamma_{F^{k(\text{gr})}}(\emptyset)$, i.e. the set of unattacked arguments does not change. Since $S \subseteq \Gamma_F(\emptyset)$ is assumed we are done. Assume now that the assertion is proven for any k -partition. Let S be a $(k+1)$ -partition, i.e. $S = \bigcup_{i=1}^{k+1} A_i$. According to induction hypothesis as well as Corollary 6.8 it suffices to prove $\bigcup_{i=1}^k A_i$ defends A_{k+1} in $F^{k(\text{gr})}$. Assume not, i.e. there are arguments $b \in A(F) \setminus S$, $c \in A_{k+1}$ s.t. $(b, c) \in R(F^{k(\text{gr})}) \subseteq R(F)$ and for all $a \in \bigcup_{i=1}^k A_i$, $(a, b) \notin R(F^{k(\text{gr})})$ (*). Since $\bigcup_{i=1}^k A_i$ defends A_{k+1} in F we deduce the existence of an argument $a \in \bigcup_{i=1}^k A_i$ s.t. $(a, b) \in R(F)$. Thus, (a, b) is redundant w.r.t. the grounded kernel. According to Definition 5.10 and due to the conflict-freeness of $\bigcup_{i=1}^k A_i$ we have $(a, a) \notin R(F)$ and $(b, a), (b, b) \in R(F)$. Consequently, $(b, a) \in F^{k(\text{gr})}$. Since $\bigcup_{i=1}^k A_i$ is a strong admissible k -partition in F we obtain by induction hypothesis that $\bigcup_{i=1}^k A_i$ is strongly admissible in $F^{k(\text{gr})}$ and therefore, admissible in $F^{k(\text{gr})}$ (Proposition 6.4). Hence there has to be an argument $a \in \bigcup_{i=1}^k A_i$ s.t. $(a, b) \in R(F^{k(\text{gr})})$, contradicting (*).

(\supseteq) Assume $S \in \text{sad}(F^{k(\text{gr})})$. We show $S \in \text{sad}(F)$ by induction on n indicating that S is a n -partition in $F^{k(\text{gr})}$. Due to $\Gamma_F(\emptyset) = \Gamma_{F^{k(\text{gr})}}(\emptyset)$ the base case is immediately clear. For the induction step let S be a $(k+1)$ -partition, i.e. $S = \bigcup_{i=1}^{k+1} A_i$. By induction hypothesis we may assume that $\bigcup_{i=1}^k A_i$ is strongly admissible in F . Using Corollary 6.8 it suffices to prove $\bigcup_{i=1}^k A_i$ defends A_{k+1} in F . Assume not, i.e. there are arguments $b \in A(F) \setminus S$, $c \in A_{k+1}$ s.t. $(b, c) \in R(F)$ and for all $a \in \bigcup_{i=1}^k A_i$, $(a, b) \notin R(F)$. We even have $(a, b) \notin R(F^{k(\text{gr})})$ since $R(F^{k(\text{gr})}) \subseteq R(F)$. Consequently, (b, c) has to be deleted in $F^{k(\text{gr})}$. Definition 5.10 requires $(c, c) \in R(F^{k(\text{gr})})$ contradicting the conflict-freeness of S in $F^{k(\text{gr})}$. \square

Theorem 6.10. *For any two AFs F and G , we have $F \equiv_E^{\text{sad}} G \iff F^{k(\text{gr})} = G^{k(\text{gr})}$.*

Proof. (\implies) We show the contrapositive, i.e. $F^{k(\text{gr})} \neq G^{k(\text{gr})} \implies F \not\equiv_E^{\text{sad}} G$. Assuming $F^{k(\text{gr})} \neq G^{k(\text{gr})}$ implies $F \not\equiv_E^{\text{gr}} G$ (cf. Theorem 5.14). This means, there is an AF H , s.t. $\text{gr}(F \dot{\cup} H) \neq \text{gr}(G \dot{\cup} H)$. Due to statement 3 of Proposition 6.4, we deduce $\text{sad}(F \dot{\cup} H) \neq \text{sad}(G \dot{\cup} H)$ proving $F \not\equiv_E^{\text{sad}} G$.

(\impliedby) Given $F^{k(\text{gr})} = G^{k(\text{gr})}$. Since expansion equivalence is a congruence w.r.t. $\dot{\cup}$ we obtain $(F \dot{\cup} H)^{k(\text{gr})} = (G \dot{\cup} H)^{k(\text{gr})}$ for any AF H . Consequently, $\text{sad}((F \dot{\cup} H)^{k(\text{gr})}) = \text{sad}((G \dot{\cup} H)^{k(\text{gr})})$. Due to Lemma 6.9 we deduce $\text{sad}(F \dot{\cup} H) = \text{sad}(G \dot{\cup} H)$, concluding the proof. \square

6.2 Verifiability

In this section we study the question whether we really need the entire AF F to compute the extensions of a given semantics. Consider naive semantics. Obviously, in order to determine naive extensions it suffices to know all conflict-free sets. Conversely, knowing $cf(F)$ only does not allow to reconstruct F unambiguously. This means, knowledge about $cf(F)$ is indeed less information than the entire AF by itself. In fact, most of the existing semantics do not need information about the entire AF. We will categorize the amount of information by taking the conflict-free sets as a basis and distinguish between different amounts of knowledge about the neighborhood of these sets. On the one hand this is natural since conflict-freeness is the most basic concept used by argumentation semantics and on the other hand the neighborhood (i.e. arguments attacking and being attacked by an argument)

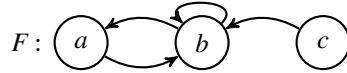
Definition 6.11. We call a function $\tau^x : 2^{\mathcal{U}} \times 2^{\mathcal{U}} \rightarrow (2^{\mathcal{U}})^n$ ($n > 0$), which is expressible via basic set operations only, *neighborhood function*. A neighborhood function τ^x induces the *verification class* mapping each AF F to $\tilde{F}^x = \{(S, \tau^x(S_F^\oplus, S_F^\ominus)) \mid S \in cf(F)\}$.

We coined the term neighborhood function because the induced verification classes apply these functions to the neighborhoods, i.e. range and anti-range of conflict-free sets. The notion of *expressible via basic set operations* simply means that (in case of $n = 1$) the expression $\tau^x(A, B)$ is in the language generated by the BNF $X ::= A \mid B \mid (X \cup X) \mid (X \cap X) \mid (X \setminus X)$. Consequently, in case of $n = 1$, we may distinguish eight set theoretically different neighborhood functions, namely

$$\begin{array}{llll} \tau^e(S, S') = \emptyset & \tau^+(S, S') = S & \tau^-(S, S') = S' & \tau^\pm(S, S') = S' \setminus S \\ \tau^\pm(S, S') = S \setminus S' & \tau^\cap(S, S') = S \cap S' & \tau^\cup(S, S') = S \cup S' & \tau^\Delta(S, S') = (S \cup S') \setminus (S \cap S') \end{array}$$

The names of the neighborhood functions are inspired by their usage in the verification classes they induce (cf. Definition 6.11). A verification class encapsulates a certain amount of information about an AF, as the following example illustrates.

Example 6.12. Consider the following AF $F = (\{a, b, c\}, \{(a, b), (b, a), (b, b), (c, b)\})$.



Now take, for instance, the verification class induced by τ^+ , that is $\tilde{F}^+ = \{(S, \tau^+(S_F^\oplus, S_F^\ominus)) \mid S \in cf(F)\} = \{(S, S_F^\oplus) \mid S \in cf(F)\}$, storing information about conflict-free sets together with their associated ranges w.r.t. F . It contains the following tuples: (\emptyset, \emptyset) , $(\{a\}, \{a, b\})$, $(\{c\}, \{b, c\})$, and $(\{a, c\}, \{a, b, c\})$. For the verification class induced by τ^\pm , on the other hand, we have $\tilde{F}^\pm = \{(\emptyset, \emptyset), (\{a\}, \emptyset), (\{c\}, \{b\}), (\{a, c\}, \emptyset)\}$.

Intuitively, it should be clear that the set \tilde{F}^+ suffices to compute stage extensions (i.e., range-maximal conflict-free sets) of F . This intuitive understanding of *verifiability* will be formally specified in Definition 6.15. Note that a neighborhood function τ^x may return n -tuples. Consequently, in consideration of the eight basic functions we obtain (modulo reordering, duplicates, empty set) $2^7 + 1$ syntactically different neighborhood functions and therefore the same number of verification classes. As usual, we denote the n -ary combination of basic functions $(\tau^{x_1}(S, S'), \dots, \tau^{x_n}(S, S'))$ as $\tau^x(S, S')$ by $x = x_1 \dots x_n$.

With the following definition we can put neighborhood functions into relation w.r.t. their information. This will help us to show that actually many of the induced classes collapse to the same amount of information.

Definition 6.13. Given neighborhood functions τ^x and τ^y returning n -tuples and m -tuples, respectively, we say that τ^x is *more informative* than τ^y , for short $\tau^x \succeq \tau^y$, iff there is a function $\delta : (2^{\mathcal{U}})^n \rightarrow (2^{\mathcal{U}})^m$ such that for any two sets of arguments $S, S' \subseteq \mathcal{U}$, we have $\delta(\tau^x(S, S')) = \tau^y(S, S')$. We denote the strict part of \succeq by \succ , i.e. $\tau^x \succ \tau^y$ iff $\tau^x \succeq \tau^y$ and $\tau^y \not\succeq \tau^x$. Finally, $\tau^x \approx \tau^y$ (τ^x *represents* τ^y and vice versa) in case $\tau^x \succeq \tau^y$ and $\tau^y \succeq \tau^x$.

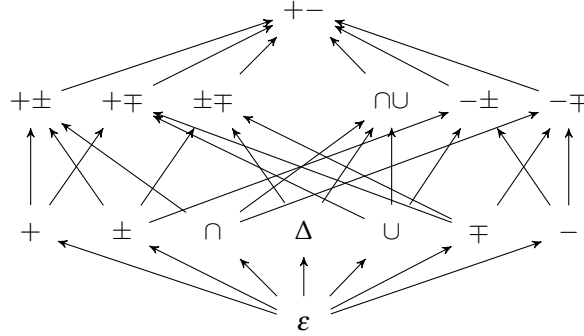


Figure 6.1: Representatives of neighborhood functions and their relation w.r.t. information; a node x stands for the neighborhood function τ^x ; an arrow from x to y means $\tau^x \prec \tau^y$.

It turns out that many neighborhood functions amount to the same amount of information. In particular, τ^{+-} represents all τ^{x_1, \dots, x_n} with $n > 2$.

Lemma 6.14. *All neighborhood functions are represented by the ones depicted in Figure 6.1 and the \prec -relation represented by arcs in Figure 6.1 holds.*

Proof. We begin by showing that all neighborhood functions are represented in Figure 6.1. Clearly, each neighborhood function τ^x represents itself, i.e. $\tau^x \approx \tau^x$. All neighborhood functions for $n = 1$ are depicted in Figure 6.1. We turn to $n = 2$. Consider the neighborhood functions $\tau^{+\pm}$, $\tau^{+\cap}$, and $\tau^{\pm\cap}$, defined as $\tau^{+\pm}(S, S') = (S, S \setminus S')$, $\tau^{+\cap}(S, S') = (S, S \cap S')$, and $\tau^{\pm\cap}(S, S') = (S \setminus S', S \cap S')$ for $S, S' \subseteq \mathcal{U}$. Observe that $S = (S \setminus S') \cup (S \cap S')$. Hence, we can easily define functions in the spirit of Definition 6.13 mapping the images of the function to one another.

- $\delta_1(\tau^{+\pm}(S, S')) = \delta_1(S, S \setminus S') =_{\text{def}} (S, S \setminus (S \setminus S')) = (S, S \cap S') = \tau^{+\cap}(S, S')$;
- $\delta_2(\tau^{+\cap}(S, S')) = \delta_2(S, S \cap S') =_{\text{def}} (S \setminus (S \cap S'), S \cap S') = (S \setminus S', S \cap S') = \tau^{\pm\cap}(S, S')$;
- $\delta_3(\tau^{\pm\cap}(S, S')) = \delta_3(S \setminus S', S \cap S') =_{\text{def}} ((S \setminus S') \cup (S \cap S'), S \setminus S') = (S, S \setminus S') = \tau^{+\pm}(S, S')$.

Therefore, $\tau^{+\pm} \approx \tau^{+\cap} \approx \tau^{\pm\cap}$. In particular, they are all represented by τ^{\pm} . We can apply the same reasoning to other combinations of neighborhood functions and get the following equivalences w.r.t. information content:

- $\tau^{+\mp} \approx \tau^{+\cup} \approx \tau^{\mp\cup}$;
- $\tau^{\pm\mp} \approx \tau^{\pm\Delta} \approx \tau^{\mp\Delta}$;
- $\tau^{\cap\cup} \approx \tau^{\cap\Delta} \approx \tau^{\cup\Delta}$;
- $\tau^{-\pm} \approx \tau^{-\cup} \approx \tau^{\pm\cup}$; and
- $\tau^{-\mp} \approx \tau^{-\cap} \approx \tau^{\mp\cap}$,

with the functions stated first acting as representatives in Figure 6.1.

For the remaining functions returning 2-tuples we get $\tau^{+-} \approx \tau^{+\Delta} \approx \tau^{-\Delta}$ by the fact that $\tau^{\Delta}(S, S') = (S \cup S') \setminus (S \cap S')$.

- $\delta_4(\tau^{+-}(S, S')) = \delta_4(S, S') =_{def} (S, (S \cup S') \setminus (S \cap S')) = \tau^{+\Delta}(S, S')$;
- $\delta_5(\tau^{+\Delta}(S, S')) = \delta_5(S, (S \cup S') \setminus (S \cap S')) =_{def} ((S \setminus ((S \cup S') \setminus (S \cap S'))) \cup ((S \cup S') \setminus (S \cap S'))) \setminus S, (S \cup S') \setminus (S \cap S')) = (S', (S \cup S') \setminus (S \cap S')) = \tau^{-\cap}(S, S')$;
- $\delta_6(\tau^{-\Delta}(S, S')) = \delta_6(S', (S \cup S') \setminus (S \cap S')) =_{def} ((S' \setminus ((S \cup S') \setminus (S \cap S'))) \cup ((S \cup S') \setminus (S \cap S'))) \setminus S', S') = (S, S') = \tau^{+-}(S, S')$.

Finally, every neighborhood function $\tau^{x_1 \dots x_n}$ with $x_1, \dots, x_n \in \{+, -, \pm, \mp, \cup, \cap, \Delta\}$ and $n \geq 3$ is represented by τ^{+-} since we can compute all possible sets from S and S' . Therefore τ^{+-} (together with all of these equally informative function) is the most informative neighborhood function.

Now consider two functions τ^x and τ^y such that there is an arrow from x to y in Figure 6.1. It is easy to see that $\tau^y \succeq \tau^x$ since, for sets of arguments S and S' , $\tau^x(S, S')$ is either contained in $\tau^y(S, S')$ or obtainable from $\tau^y(S, S')$ by basic set operations. The fact that $\tau^x \not\succeq \tau^y$, entailing $\tau^y \succ \tau^x$, follows from the impossibility of finding a function δ such that $\delta(\tau^x(S, S')) = \tau^y(S, S')$. \square

If the information provided by a neighborhood function is sufficient to compute the extensions under a semantics, we say that the semantics is verifiable by the class induced by the neighborhood function.

Definition 6.15. A semantics σ is *verifiable* by the verification class induced by the neighborhood function τ^x returning n -tuples (or simply, *x-verifiable*) iff there is a function (also called *criterion*) $\gamma_\sigma : (2^{\mathcal{U}})^n \times 2^{\mathcal{U}} \rightarrow 2^{\mathcal{U}}$ s.t. for every AF $F \in \mathcal{U}$ we have: $\gamma_\sigma(\tilde{F}^x, A(F)) = \sigma(F)$. Moreover, σ is *exactly x-verifiable* iff σ is x -verifiable and there is no verification class induced by τ^y with $\tau^y \prec \tau^x$ such that σ is y -verifiable.

Observe that if a semantics σ is x -verifiable then for any two AFs F and G with $\tilde{F}^x = \tilde{G}^x$ and $A(F) = A(G)$ it must hold that $\sigma(F) = \sigma(G)$.

We proceed with a list of criteria showing that well-known semantics mentioned in Definition 3.4 are verifiable by a verification class induced by a certain neighborhood function. In the following, we abbreviate the tuple $(\tilde{F}^x, A(F))$ by \tilde{F}_A^x .

$$\begin{aligned}
 \gamma_{na}(\tilde{F}_A^e) &= \{S \mid S \in \tilde{F}, S \text{ is } \subseteq\text{-maximal in } \tilde{F}\}; \\
 \gamma_{stg}(\tilde{F}_A^+) &= \{S \mid (S, S^+) \in \tilde{F}^+, S^+ \text{ is } \subseteq\text{-maximal in } \{C^+ \mid (C, C^+) \in \tilde{F}^+\}\}; \\
 \gamma_{stb}(\tilde{F}_A^+) &= \{S \mid (S, S^+) \in \tilde{F}^+, S^+ = A\}; \\
 \gamma_{ad}(\tilde{F}_A^\mp) &= \{S \mid (S, S^\mp) \in \tilde{F}^\mp, S^\mp = \emptyset\}; \\
 \gamma_{pr}(\tilde{F}_A^\mp) &= \{S \mid S \in \gamma_{ad}(\tilde{F}_A^\mp), S \text{ is } \subseteq\text{-maximal in } \gamma_{ad}(\tilde{F}_A^\mp)\}; \\
 \gamma_{ss}(\tilde{F}_A^{+\mp}) &= \{S \mid S \in \gamma_{ad}(\tilde{F}_A^\mp), S^+ \text{ is } \subseteq\text{-maximal in } \{C^+ \mid (C, C^+, C^\mp) \in \tilde{F}^{+\mp}, C \in \gamma_{ad}(\tilde{F}_A^\mp)\}\}; \\
 \gamma_{il}(\tilde{F}_A^\mp) &= \{S \mid S \text{ is } \subseteq\text{-maximal in } \{C \mid C \in \gamma_{ad}(\tilde{F}_A^\mp), C \subseteq \bigcap \gamma_{pr}(\tilde{F}_A^\mp)\}\}; \\
 \gamma_{eg}(\tilde{F}_A^{+\mp}) &= \{S \mid S \text{ is } \subseteq\text{-maximal in } \{C \mid C \in \gamma_{ad}(\tilde{F}_A^\mp), C \subseteq \bigcap \gamma_{ss}(\tilde{F}_A^{+\mp})\}\};
 \end{aligned}$$

$$\begin{aligned}
\gamma_{sad}(\tilde{F}_A^{\pm}) &= \left\{ S \mid (S, S^-, S^\pm) \in \tilde{F}^{\pm}, \exists (S_0, S_0^-, S_0^\pm), \dots, (S_n, S_n^-, S_n^\pm) \in \tilde{F}^{\pm} : \right. \\
&\quad \left. (\emptyset = S_0 \subset \dots \subset S_n = S \wedge \forall i \in \{1, \dots, n\} : S_i^- \subseteq S_{i-1}^\pm) \right\}; \\
\gamma_{gr}(\tilde{F}_A^{\pm}) &= \left\{ S \mid S \in \gamma_{sad}(\tilde{F}_A^{\pm}), \forall (\bar{S}, \bar{S}^-, \bar{S}^\pm) \in \tilde{F}^{\pm} : \bar{S} \supset S \Rightarrow (\bar{S}^- \setminus S^\pm) \neq \emptyset \right\}; \\
\gamma_{co}(\tilde{F}_A^{\pm}) &= \left\{ S \mid (S, S^+, S^-) \in \tilde{F}^{\pm}, (S^- \setminus S^+) = \emptyset, \forall (\bar{S}, \bar{S}^+, \bar{S}^-) \in \tilde{F}^{\pm} : \bar{S} \supset S \Rightarrow (\bar{S}^- \setminus S^+) \neq \emptyset \right\}.
\end{aligned}$$

It is easy to see that the naive semantics is verifiable by the verification class induced by τ^e since the naive extensions can be determined by the conflict-free sets. Stable and stage semantics, on the other hand, utilize the range of each conflict-free set in addition. Hence they are verifiable by the verification class induced by τ^+ . Now consider admissible sets. Recall that a conflict-free S set is admissible if and only if it attacks all attackers. This is captured exactly by the condition $S^\mp = \emptyset$, hence admissible sets are verifiable by the verification class induced by τ^\mp . The same holds for preferred semantics, since we just have to determine the maximal conflict-free sets with $S^\mp = \emptyset$. Semi-stable semantics, however, needs the range of each conflict-free set in addition, see γ_{ss} , which makes it verifiable by the verification class induced by $\tau^{\mp\pm}$. Finally consider the criterion γ_{co} . The first two conditions for a set of arguments S stand for conflict-freeness and admissibility, respectively. Now assume the third condition does not hold, i.e., there exists a tuple $(\bar{S}, \bar{S}^+, \bar{S}^-) \in \tilde{F}^{\pm}$ with $\bar{S} \supset S$ and $\bar{S}^- \setminus S^+ = \emptyset$. This means that every argument attacking \bar{S} is attacked by S , i.e., \bar{S} is defended by S . Hence S is not a complete extension, showing that $\gamma_{co}(\tilde{F}_A^{\pm}) = co(F)$ for each $F \in \mathcal{U}$. One can verify that all criteria from the list are adequate in the sense that they describe the extensions of the corresponding semantics.

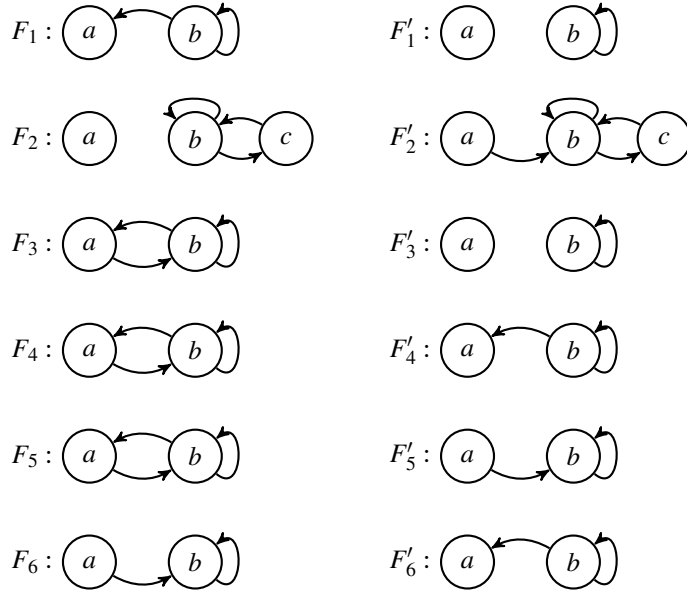
The concepts of verifiability and being more informative behave correctly insofar as more informative neighborhood functions do not lead to a loss of verification capacity.

Proposition 6.16. *If a semantics σ is x -verifiable, then σ is verifiable by all verification classes induced by some τ^y with $\tau^y \succeq \tau^x$.*

Proof. As σ is verifiable by the verification class induced by τ^x it holds that there is some γ_σ such that for all $F \in \mathcal{U}$, $\gamma_\sigma(\tilde{F}^x, A(F)) = \sigma(F)$. Now let $\tau^y \succeq \tau^x$, meaning that there is some δ such that $\delta(\tau^y(S, S')) = \tau^x$. We define $\gamma'_\sigma(\tilde{F}^y, A(F)) = \gamma_\sigma(\{(S, \delta(S)) \mid (S, S) \in \tilde{F}^y\}, A(F))$ and observe that $\{(S, \delta(S)) \mid (S, S) \in \tilde{F}^y\} = \tilde{F}^x$, hence $\gamma'_\sigma(\tilde{F}^y, A(F)) = \sigma(F)$ for each $F \in \mathcal{U}$. \square

In order to prove unverifiability of a semantics σ w.r.t. a class induced by a certain τ^x it suffices to present two AFs F and G such that $\sigma(F) \neq \sigma(G)$ but, $\tilde{F}^x = \tilde{G}^x$ and $A(F) = A(G)$. Then the verification class induced by τ^x does not provide enough information to verify σ . In the following we will use this strategy to show exact verifiability. Consider a semantics σ which is verifiable by a class induced by τ^x . If σ is unverifiable by all verifiability classes induced by τ^y with $\tau^y \prec \tau^x$ we have that σ is exactly verifiable by τ^x . The following examples study this issue for the semantics under consideration.

Example 6.17. The complete semantics is $+-$ -verifiable as seen before. The following AFs show that it is even exactly verifiable by that class.



First consider the AFs F_1 and F'_1 , and observe that $\widetilde{F_1}^{+\pm} = \{(\emptyset, \emptyset, \emptyset), (\{a\}, \emptyset, \emptyset)\} = \widetilde{F'_1}^{+\pm}$. On the other hand F_1 and F'_1 differ in their complete extensions since $co(F_1) = \{\emptyset\}$ but $co(F'_1) = \{\{a\}\}$. Therefore complete semantics is unverifiable by the verification class induced by $\tau^{+\pm}$. Likewise, this can be shown for the classes induced by $\tau^{-\mp}$, $\tau^{\pm\mp}$, $\tau^{-\pm}$, $\tau^{+\mp}$, and $\tau^{\cap\cup}$, respectively:

- $\widetilde{F_2}^{-\mp} = \widetilde{F'_2}^{-\mp}$, but $co(F_2) = \{\{a\}, \{a, c\}\} \neq \{\{a, c\}\} = co(F'_2)$.
- $\widetilde{F_3}^{\pm\mp} = \widetilde{F'_3}^{\pm\mp}$, but $co(F_3) = \{\emptyset, \{a\}\} \neq \{\{a\}\} = co(F'_3)$.
- $\widetilde{F_4}^{-\pm} = \widetilde{F'_4}^{-\pm}$, but $co(F_4) = \{\emptyset, \{a\}\} \neq \{\emptyset\} = co(F'_4)$.
- $\widetilde{F_5}^{+\mp} = \widetilde{F'_5}^{+\mp}$, but $co(F_5) = \{\emptyset, \{a\}\} \neq \{\{a\}\} = co(F'_5)$.
- $\widetilde{F_6}^{\cap\cup} = \widetilde{F'_6}^{\cap\cup}$, but $co(F_6) = \{\{a\}\} \neq \{\emptyset\} = co(F'_6)$.

Hence complete semantics is exactly verifiable by the verification class induced by τ^{+-} .

Example 6.18. Consider the semi-stable and eager semantics and recall that they are $+-$ -verifiable. To show exact verifiability it suffices to show unverifiability by the classes induced by τ^+ , τ^{\cup} , and τ^{\mp} (cf. Figure 6.1); F_1 and F_6 are taken from Example 6.17 above.

- $\widetilde{F_1}^+ = \widetilde{F'_1}^+$, but $ss(F_1) = eg(F_1) = \{\emptyset\} \neq \{\{a\}\} = ss(F'_1) = eg(F'_1)$.
- $\widetilde{F_6}^{\cup} = \widetilde{F'_6}^{\cup}$, but $ss(F_6) = eg(F_6) = \{\{a\}\} \neq \{\emptyset\} = ss(F'_6) = eg(F'_6)$.
- $\widetilde{F_7}^{\mp} = \widetilde{F'_7}^{\mp}$, but $ss(F_7) = \{\{b\}\} \neq \{\{a\}, \{b\}\} = ss(F'_7)$ and $eg(F_7) = \{\{b\}\} \neq \{\emptyset\} = eg(F'_7)$.



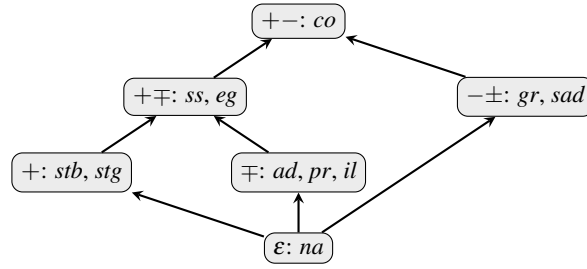


Figure 6.2: Semantics and their Exact Verification Classes

Hence, both the semi-stable and eager semantics are exactly verifiable by the verification class induced by $\tau^{+\mp}$.

Example 6.19. Now consider the grounded and strong admissible semantics and recall that they are $-\pm$ -verifiable. In order to show exact verifiability we have to show unverifiability by the classes induced by τ^{\pm} , τ^{-} , and τ^{\cup} (cf. Figure 6.1); again, the AFs from Example 6.17 can be reused.

- $\widetilde{F}_1^{\pm} = \widetilde{F}_1'^{\pm}$, but $gr(F_1) = \{\emptyset\} \neq \{\{a\}\} = gr(F_1')$ and $sad(F_1) = \{\emptyset\} \neq \{\emptyset, \{a\}\} = sad(F_1')$.
- $\widetilde{F}_2^{-} = \widetilde{F}_2'^{-}$, but $gr(F_2) = \{\{a\}\} \neq \{\{a, c\}\} = gr(F_2')$ and $sad(F_2) = \{\emptyset, \{a\}\} \neq \{\emptyset, \{a\}, \{a, c\}\} = sad(F_2')$.
- $\widetilde{F}_6^{\cup} = \widetilde{F}_6'^{\cup}$, but $gr(F_6) = \{\{a\}\} \neq \{\emptyset\} = gr(F_6')$ and $sad(F_6) = \{\emptyset, \{a\}\} \neq \{\emptyset\} = sad(F_6')$.

Hence, both the grounded and strong admissible semantics are exactly verifiable by the verification class induced by $\tau^{+\mp}$.

Example 6.20. Finally consider stable, stage, admissible, preferred and ideal semantics. They are either $+$ -verifiable (stb and stg) or \mp -verifiable (ad , pr , and il). We have to show unverifiability w.r.t. the verification class induced by τ^{ε} . Consider, for instance, the AFs F_4 and F_4' from Example 6.17. We have $\widetilde{F}_4^{\varepsilon} = \widetilde{F}_4'^{\varepsilon}$, but $ad(F_4) = \{\emptyset, \{a\}\} \neq \{\emptyset\} = ad(F_4')$, $stb(F_4) = \{\{a\}\} \neq \emptyset = stb(F_4')$, and $\sigma(F_4) = \{\{a\}\} \neq \{\emptyset\} = \sigma(F_4')$ for $\sigma \in \{stg, pr, il\}$, showing exactness of the respective verification classes.

The insights obtained through Examples 6.17, 6.18, 6.19 and 6.20 show that the verification classes obtained from the criteria given above are indeed exact. Figure 6.2 shows the resulting relation between the semantics under consideration with respect to their exact verification classes.

We turn now to the main theorem stating that any *rational* semantics is exactly verifiable by one of the 15 different verification classes.

Theorem 6.21. *Every semantics which is rational is exactly verifiable by a verification class induced by one of the neighborhood functions presented in Figure 6.1.*

Proof. First of all note that by Lemma 6.14, τ^{ε} is the least informative neighborhood function and for every other neighborhood function τ^x it holds that $\tau^{\varepsilon} \preceq \tau^x$. Therefore, if a semantics is verifiable by the verification class induced by any τ^x then it is exactly verifiable by a verification class induced by some τ^y with $\tau^{\varepsilon} \preceq \tau^y \preceq \tau^x$. Moreover, if a semantics is exactly verifiable by a class, then it is by definition also verifiable by this class. Hence it remains to show that every semantics which is rational is verifiable by a verification class presented in Figure 6.1.

We show the contrapositive, i.e., if a semantics is not verifiable by a verification class induced by one of the neighborhood functions presented in Figure 6.1 then it is not rational. Assume a semantics

σ is not verifiable by one of the verification classes. This means σ is not verifiable by the verification class induced by τ^{+-} . Hence there exist two AFs F and G such that $\tilde{F}^{+-} = \tilde{G}^{+-}$ and $A(F) = A(G)$, but $\sigma(F) \neq \sigma(G)$. For every argument a which is not self-attacking, a tuple $(\{a\}, \{a\}^+, \{a\}^-)$ is contained in \tilde{F}^{+-} (and in \tilde{G}^{+-}). Hence F and G have the same not-self-attacking arguments and, moreover these arguments have the same ingoing and outgoing attacks in F and G . This, together with $A(F) = A(G)$ implies that $F^l = G^l$ (see Definition 6.2) holds. But since $\sigma(F) \neq \sigma(G)$ we get that σ is not rational, which was to show. \square

Note that the criterion giving evidence for verifiability of a semantics by a certain class has access to the set of arguments of a given AF. In fact, only the criterion for stable semantics makes use of that – it can be omitted for the other semantics.

6.3 Intermediate Semantics

A type of semantics which has aroused quite some interest in the literature (see e.g. [Baroni and Giacomin, 2007a] and [Nieves et al., 2011]) are intermediate semantics, i.e. semantics which yield results lying between two existing semantics. The introduction of σ - τ -intermediate semantics can be motivated by deleting *undesired* (or add *desired*) τ -extensions² while guaranteeing all reasonable positions w.r.t. σ . In other words, σ - τ -intermediate semantics can be seen as sceptical or credulous acceptance shifts within the range of σ and τ .

A natural question is whether we can make any statements about compatible kernels of intermediate semantics. In particular, if semantics σ and τ are characterizable by some kernel k , is then every σ - τ -intermediate semantics characterizable by k . The following example answers this question negatively.

Example 6.22. Recall from Theorem 5.14 that both stable and stage semantics are compatible with $k(stb)$, i.e. $F \equiv_E^{stb} G \iff F \equiv_E^{stg} G \iff F^{k(stb)} = G^{k(stb)}$. Now we define the following *stb-stg*-intermediate semantics, say *stagle* semantics: Given an AF $F = (A, R)$, $S \in sta(F)$ iff $S \in cf(F)$, $S_F^\oplus \cup S_F^\ominus = A$ and for every $T \in cf(F)$ we have $S_F^\oplus \not\subseteq T_F^\oplus$. Obviously, it holds that $stb \subseteq sta \subseteq stg$ and $stb \neq sta$ as well as $sta \neq stg$, as witnessed by the AF F :



It is easy to verify that $stb(F) = \emptyset \subset sta(F) = \{\{b\}\} \subset stg(F) = \{\{b\}, \{c\}\}$. We proceed by showing that *stagle* semantics is not characterizable by $k(stb)$. To this end consider $F^{k(stb)}$ which is depicted below.



Now, $sta(F^{k(stb)}) = \{\{b\}, \{c\}\}$ witnesses $F \not\equiv_E^{sta} F^{k(stb)}$ and therefore, $F \not\equiv_E^{stb} F^{k(stb)}$. Since $F^{k(stb)} = (F^{k(stb)})^{k(stb)}$ we are done, i.e. *stagle* semantics is indeed not characterizable by the stable kernel.

It is the main result of this section that characterizability of intermediate semantics w.r.t. a certain kernel can be guaranteed if verifiability w.r.t. a certain class is presumed. The provided characterization theorems generalize former results presented in Theorem 5.14. Moreover, due to the abstract character of the theorems the results are applicable to semantics which may be defined in the future.

²Recently, the so-called *extension removal problem* was studied []. Here, instead of sticking to an intermediate semantics the authors studied whether it is possible - and if so how - to modify a given AF in such a way that certain undesired extensions are no longer generated.

Before turning to the characterization theorems we state some implications of verifiability. In particular, under the assumption that σ is verifiable by a certain class, equality of certain kernels implies expansion equivalence w.r.t. σ .

Proposition 6.23. *For any +-verifiable semantics σ we have $F^{k(stb)} = G^{k(stb)} \implies F \equiv_E^\sigma G$.*

Proof. It was shown that $F^{k(stb)} = G^{k(stb)} \implies (F \dot{\cup} H)^{k(stb)} = (G \dot{\cup} H)^{k(stb)}$ (cf. In [Oikarinen and Woltran, 2011] or Item 5 of Fact 5.12(i)). Consider now a +-verifiable semantics σ . In order to show $\sigma(F) = \sigma(G^{k(stb)})$ (ii) we prove $\widetilde{F}^+ = \widetilde{F^{k(stb)}}^+$ (*) first. It is easy to see that $S \in cf(F)$ iff $S \in cf(G^{k(stb)})$. Furthermore, since $k(stb)$ deletes an attack (a, b) only if a is self-defeating we deduce that ranges does not change as long as conflict-free sets are considered. Thus, $\sigma(F) =_{(Def.)} \gamma_\sigma(\widetilde{F}^+) =_{(*)} \gamma_\sigma(\widetilde{F^{k(stb)}}^+) =_{(Def.)} \sigma(G^{k(stb)})$. Now assume that $F^{k(stb)} = G^{k(stb)}$ and let $S \in \sigma(F \dot{\cup} H)$ for some AF H . We have to show that $S \in \sigma(G \dot{\cup} H)$. Applying (ii) we obtain $S \in \sigma((F \dot{\cup} H)^{k(stb)})$. Furthermore, using (i) we deduce $S \in \sigma((G \dot{\cup} H)^{k(stb)})$. Finally, $S \in \sigma(G \dot{\cup} H)$ by applying (ii), which concludes the proof. \square

The following results can be shown in a similar manner.

Proposition 6.24. *For a semantics σ it holds that*

- if σ is +-verifiable then $F^{k(stb)} = G^{k(stb)} \implies F \equiv_E^\sigma G$.
- if σ is + \mp -verifiable then $F^{k(ad)} = G^{k(ad)} \implies F \equiv_E^\sigma G$.
- if σ is +- -verifiable then $F^{k(co)} = G^{k(co)} \implies F \equiv_E^\sigma G$.
- if σ is - \pm -verifiable then $F^{k(gr)} = G^{k(gr)} \implies F \equiv_E^\sigma G$.
- if σ is ε -verifiable then $F^{k(na)} = G^{k(na)} \implies F \equiv_E^\sigma G$.

We proceed with general characterization theorems. The first one states that *stb-stg*-intermediate semantics are characterizable via the stable kernel if +-verifiability is given. Consequently, stale semantics as defined in Example 6.22 can not be +-verifiable.

Theorem 6.25. *Given a semantics σ which is +-verifiable and stb-stg-intermediate, then $F^{k(stb)} = G^{k(stb)} \iff F \equiv_E^\sigma G$.*

Proof. (\implies) Follows directly from Proposition 6.24.

(\impliedby) We show the contrapositive, i.e. $F^{k(stb)} \neq G^{k(stb)} \implies F \not\equiv_E^\sigma G$. Assuming $F^{k(stb)} \neq G^{k(stb)}$ implies $F \not\equiv_E^{stb} G$, i.e. there exists an AF H such that $stb(F \dot{\cup} H) \neq stb(G \dot{\cup} H)$ and therefore, $stb(F \dot{\cup} H) \neq stb(G \dot{\cup} H)$. Let $B = A(F) \cup A(G) \cup A(H)$ and $H' = (B \cup \{a\}, \{(a, b), (b, a) \mid b \in B\})$. It is easy to see that $stb(F \dot{\cup} H') = stb(F \dot{\cup} H) \cup \{\{a\}\}$ and $stb(G \dot{\cup} H') = stb(G \dot{\cup} H) \cup \{\{a\}\}$. Since now both $stb(F \dot{\cup} H') \neq \emptyset$ and $stb(G \dot{\cup} H') \neq \emptyset$ it holds that $stb(F \dot{\cup} H') = stb(F \dot{\cup} H')$ and $stb(G \dot{\cup} H') = stb(G \dot{\cup} H')$. Hence $\sigma(F \dot{\cup} H') \neq \sigma(G \dot{\cup} H')$, showing that $F \not\equiv_E^{stb} G$. \square

The following theorems can be shown in a similar manner.

Theorem 6.26. *Given a semantics σ which is - \pm -verifiable and gr-sad-intermediate, it holds that $F^{k(gr)} = G^{k(gr)} \iff F \equiv_E^\sigma G$.*

Theorem 6.27. *Given a semantics σ which is $+\mp$ -verifiable and ρ -ad-intermediate for any $\rho \in \{ss, il, eg\}$, it holds that $F^{k(ad)} = G^{k(ad)} \iff F \equiv_E^\sigma G$.*

Recall that complete semantics is a *ss-ad*-intermediate semantics. Furthermore, it is not characterizable by the admissible kernel as already observed in [Oikarinen and Woltran, 2011]. Consequently, it is not $+\mp$ -verifiable (as we have shown in Example 6.17 with considerable effort).

6.4 Conclusions

In this chapter we have initiated a, to the best of our knowledge, novel approach contributed to the analysis and comparison of abstract argumentation semantics. The main idea of our approach is to provide a novel categorization in terms of the amount of information required for testing whether a set of arguments is an extension of a certain semantics. The resulting notion of verification classes allows us to categorize any new semantics (given it is “rational”) with respect to the information needed and compare it to other semantics. Thus our work is in the tradition of the principle-based evaluation due to Baroni and Giacomin [Baroni and Giacomin, 2007b] and paves the way for a more general view on argumentation semantics, their common features, and their inherent differences.

Using our notion of verifiability, we were able to show characterizability for certain intermediate semantics w.r.t. some classical kernels. Concerning concrete semantics, our results yield the following observation: While preferred, semi-stable, ideal and eager semantics coincide w.r.t. strong equivalence, verifiability of these semantics differs. In fact, preferred and ideal semantics manage to be verifiable with strictly less information.

For future work we envisage an extension of the notion of verifiability classes in order to categorize semantics not captured by the approach followed in this chapter, such as *cf2* [Baroni et al., 2005].

Chapter 7

Summary and Final Remarks

The field of Artificial Intelligence (AI) brings together people from theoretical sciences like mathematics or philosophy and application oriented researchers. For a particular application the latter group is engaged with several design decisions ranging from *how to represent the relevant knowledge?* and *how to update a given knowledge base?* over *how to simplify a given knowledge base?* to *how much computing power is sufficient for the intelligent agent?*. In consideration of the large variety of existing logical formalisms it is of utmost importance to select the most adequate one for the specific purpose in mind. The presented habilitation treatise tackles several fundamental intrinsic properties of knowledge representation formalisms and thus contributes to an informed choice for the designer. The results can be compactly summarized as follows (for detailed summaries and conclusions we refer to the Sections 2.4, 3.5, 4.7, 5.5 and 6.4). First, we tackled the open question whether strong equivalence of two theories (which guarantees their mutual replaceability without loss of information) in a certain formalism can be decided via ordinary equivalence in another classical logic-like formalism. We showed that the important case of considering only finite knowledge bases guarantees the existence of a canonical characterizing formalism. This means that the search for characterizing logics for a given representation formalism – analogously to the logic of here-and-there in case of normal logic programs under stable model semantics [Lifschitz et al., 2001] – is not doomed to failure. Secondly, we studied argumentation semantics which play the flagship role in Dung’s abstract argumentation theory in very detail [Dung, 1995]. In particular, we compared a representative number of semantics regarding the following properties: existence and uniqueness, expressibility, replaceability and verifiability. It turned out that the considered properties are highly sensitive to the chosen semantics as well as structural properties of the considered AFs. This variety can be seen as a positive feature since it allows the designer to choose the most appropriate one for the current task.

In the last few years AI has stepped more and more into the public. Omniscient question answering systems like IBM’s *Watson* [Ferrucci et al., 2010] or a superhuman chess, shogi as well as Go playing program like DeepMind’s *AlphaZero* [Silver et al., 2018] left a remarkable impression in the wider public. Especially, the latter program is considered to be one of the greatest breakthroughs in AI. Even the most optimistic researchers did not expect that a game like Go will be played by a machine at super human level within the next 10 years since it is an extremely complex game, more so than chess. The Danish chess grandmaster Peter Heine Nielsen [BBC, 2017] said about the chess skills of AlphaZero the following:

I always wondered how it would be if a superior species landed on earth and showed us how they played chess. Now I know.

The public and scientific opinion about what AI is capable to achieve changed so many times since the by now famous *Dartmouth Summer Workshop* in 1956 [McCarthy et al., 1955] which is considered

by most researchers as the beginning of AI research. The history of artificial intelligence is marked by the alternation of periods of enthusiasm equipped with broad funding opportunities and cycles of disappointment and criticism, so-called *AI winters*. Right now we are definitely in a hype cycle. Several major biennial conferences in AI decided to become an annual event like the prestigious *International Joint Conference on Artificial Intelligence* (from 2015 onwards) [IJC, 2019] as well as the leading knowledge representation conference *International Conference on Principles of Knowledge Representation and Reasoning* (from 2020 onwards) [KR, 2019]. Moreover, the German government decided to spend more than 3 billion Euros for research and development of artificial intelligence over the next years up to 2025 [Scherer and Lotito, 2018]. In order to stop the *AI brain drain* the initiative will include a funding of 100 new professorships. The goal is clear: Germany wants to become a worldwide leader in AI.

I'm sure that the next few years will be an exciting time regarding new AI developments. One of the main challenges will be how to transfer research results into industry and finally into our daily life. It is not too far-fetched to envisage that the expertise of programs like AlphaZero and Watson might be used to optimize dangerous missions like fire fighting or to assist elderly people in their daily decisions. What is not achieved by both programs so far is the ability to explain why certain decisions are derived. This means, the actions performed by these AIs are neither transparent, nor easily understandable by humans. In order to attain a broad approval for AI technologies explanation components has to be implemented. This issue or this demand is known as *explainable AI* [Holzinger, 2018]. Argumentation theory provides the formal ingredients for being able to have a discussion or justifying a certain proposal via explicitly mentioning the pros and cons. In this regard, I believe that the presented formal study of intrinsic properties of Dung's abstract argumentation formalism may contribute to the development of a new AI generation.

Bibliography

- [IJC, 2019] (2019). International Joint Conferences on Artificial Intelligence Organization. <https://www.ijcai.org/>.
- [KR, 2019] (2019). Principles of Knowledge Representation and Reasoning. <http://www.kr.org/>.
- [Alchourrón et al., 1985] Alchourrón, C. E., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530.
- [Arieli, 2012] Arieli, O. (2012). Conflict-tolerant semantics for argumentation frameworks. In *Logics in Artificial Intelligence - 13th European Conference, JELIA 2012, Toulouse, France, September 26-28, 2012. Proceedings*, pages 28–40.
- [Baroni et al., 2014] Baroni, P., Boella, G., Cerutti, F., Giacomin, M., van der Torre, L. W. N., and Villata, S. (2014). On the input/output behavior of argumentation frameworks. *Artificial Intelligence*, 217:144–197.
- [Baroni et al., 2011a] Baroni, P., Caminada, M., and Giacomin, M. (2011a). An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26:365–410.
- [Baroni et al., 2018a] Baroni, P., Caminada, M., and Giacomin, M. (2018a). Abstract argumentation frameworks and their semantics. In Baroni, P., Gabbay, D., Giacomin, M., and van der Torre, L., editors, *Handbook of Formal Argumentation*, chapter 4. College Publications.
- [Baroni et al., 2013] Baroni, P., Cerutti, F., Dunne, P. E., and Giacomin, M. (2013). Automata for infinite argumentation structures. *Artificial Intelligence*, 203:104–150.
- [Baroni et al., 2011b] Baroni, P., Dunne, P. E., and Giacomin, M. (2011b). On the resolution-based family of abstract argumentation semantics and its grounded instance. *Artificial Intelligence*, 175(3-4):791–813.
- [Baroni and Giacomin, 2007a] Baroni, P. and Giacomin, M. (2007a). Comparing argumentation semantics with respect to skepticism. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 9th European Conference, ECSQARU 2007, Hammamet, Tunisia, October 31 - November 2, 2007, Proceedings*, pages 210–221.
- [Baroni and Giacomin, 2007b] Baroni, P. and Giacomin, M. (2007b). On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171:675–700.
- [Baroni et al., 2005] Baroni, P., Giacomin, M., and Guida, G. (2005). SCC-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168:162–210.

- [Baroni et al., 2018b] Baroni, P., Giacomin, M., and Liao, B. (2018b). Locality and modularity in abstract argumentation. In Baroni, P., Gabbay, D., Giacomin, M., and van der Torre, L., editors, *Handbook of Formal Argumentation*, chapter 19. College Publications.
- [Baumann, 2011] Baumann, R. (2011). Splitting an argumentation framework. In *Logic Programming and Nonmonotonic Reasoning - 11th International Conference, LPNMR 2011, Vancouver, Canada, May 16-19, 2011. Proceedings*, pages 40–53.
- [Baumann, 2012a] Baumann, R. (2012a). Normal and strong expansion equivalence for argumentation frameworks. *Artificial Intelligence*, 193:18–44.
- [Baumann, 2012b] Baumann, R. (2012b). What does it take to enforce an argument? minimal change in abstract argumentation. In *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31, 2012*, pages 127–132.
- [Baumann, 2014a] Baumann, R. (2014a). Context-free and context-sensitive kernels: Update and deletion equivalence in abstract argumentation. In *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, pages 63–68.
- [Baumann, 2014b] Baumann, R. (2014b). *Metalogical Contributions to the Nonmonotonic Theory of Abstract Argumentation*. College Publications - Studies in Logic.
- [Baumann, 2016] Baumann, R. (2016). Characterizing equivalence notions for labelling-based semantics. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016.*, pages 22–32.
- [Baumann, 2018] Baumann, R. (2018). On the nature of argumentation semantics: Existence and uniqueness, expressibility, and replaceability. In Baroni, P., Gabbay, D., Giacomin, M., and van der Torre, L., editors, *Handbook of Formal Argumentation*, chapter 14. College Publications. also appears in *IfCoLog Journal of Logics and their Applications* 4(8):2779-2886.
- [Baumann and Brewka, 2010] Baumann, R. and Brewka, G. (2010). Expanding argumentation frameworks: Enforcing and monotonicity results. In *Computational Models of Argument: Proceedings of COMMA 2010, Desenzano del Garda, Italy, September 8-10, 2010.*, pages 75–86.
- [Baumann and Brewka, 2013a] Baumann, R. and Brewka, G. (2013a). Analyzing the equivalence zoo in abstract argumentation. In *Computational Logic in Multi-Agent Systems - 14th International Workshop, CLIMA XIV, Corunna, Spain, September 16-18, 2013. Proceedings*, pages 18–33.
- [Baumann and Brewka, 2013b] Baumann, R. and Brewka, G. (2013b). Spectra in abstract argumentation: An analysis of minimal change. In *Logic Programming and Nonmonotonic Reasoning, 12th International Conference, LPNMR 2013, Corunna, Spain, September 15-19, 2013. Proceedings*, pages 174–186.
- [Baumann and Brewka, 2015] Baumann, R. and Brewka, G. (2015). AGM meets abstract argumentation: Expansion and revision for dung frameworks. In *IJCAI, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 2734–2740.
- [Baumann and Brewka, 2018] Baumann, R. and Brewka, G. (2018). The equivalence zoo for dung-style semantics. *Journal of Logic and Computation*, 28(3):477–498.

- [Baumann et al., 2012] Baumann, R., Brewka, G., Dvorák, W., and Woltran, S. (2012). Parameterized splitting: A simple modification-based approach. In *Correct Reasoning - Essays on Logic-Based AI in Honour of Vladimir Lifschitz*, pages 57–71.
- [Baumann et al., 2011] Baumann, R., Brewka, G., and Wong, R. (2011). Splitting argumentation frameworks: An empirical evaluation. In *Theorie and Applications of Formal Argumentation - First International Workshop, TAFA 2011, Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, pages 17–31.
- [Baumann et al., 2016a] Baumann, R., Dvorák, W., Linsbichler, T., Spanring, C., Strass, H., and Woltran, S. (2016a). On rejected arguments and implicit conflicts: The hidden power of argumentation semantics. *Artificial Intelligence*, 241:244–284.
- [Baumann et al., 2014a] Baumann, R., Dvorák, W., Linsbichler, T., Strass, H., and Woltran, S. (2014a). Compact argumentation frameworks. In *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, pages 69–74.
- [Baumann et al., 2014b] Baumann, R., Dvorák, W., Linsbichler, T., Strass, H., and Woltran, S. (2014b). Compact argumentation frameworks. *CoRR*, abs/1404.7734.
- [Baumann et al., 2017] Baumann, R., Dvořák, W., Linsbichler, T., and Woltran, S. (2017). A general notion of equivalence for abstract argumentation. In *IJCAI, Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, China, Australia, August 19-25, 2017*, page to appear.
- [Baumann et al., 2016b] Baumann, R., Linsbichler, T., and Woltran, S. (2016b). Verifiability of argumentation semantics. In *Computational Models of Argument - Proceedings of COMMA 2016, Potsdam, Germany, 12-16 September, 2016.*, pages 83–94.
- [Baumann et al., 2016c] Baumann, R., Linsbichler, T., and Woltran, S. (2016c). Verifiability of argumentation semantics. *CoRR*, abs/1603.09502.
- [Baumann and Spanring, 2015] Baumann, R. and Spanring, C. (2015). Infinite argumentation frameworks - On the existence and uniqueness of extensions. In *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation - Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, pages 281–295.
- [Baumann and Spanring, 2017] Baumann, R. and Spanring, C. (2017). A study of unrestricted abstract argumentation frameworks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 807–813.
- [Baumann and Strass, 2012] Baumann, R. and Strass, H. (2012). Default reasoning about actions via abstract argumentation. In *Computational Models of Argument: Proceedings of COMMA 2012., September 10-12, 2012, Vienna, Austria*, pages 297–309.
- [Baumann and Strass, 2013] Baumann, R. and Strass, H. (2013). On the maximal and average numbers of stable extensions. In *Theory and Applications of Formal Argumentation - Second International Workshop, TAFA 2013, Beijing, China, August 3-5, 2013, Revised Selected papers*, pages 111–126.
- [Baumann and Strass, 2015] Baumann, R. and Strass, H. (2015). Open problems in abstract argumentation. In *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation - Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, pages 325–339.

- [Baumann and Strass, 2016] Baumann, R. and Strass, H. (2016). An abstract logical approach to characterizing strong equivalence in logic-based knowledge representation formalisms. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016.*, pages 525–528.
- [Baumann and Strass, 2017] Baumann, R. and Strass, H. (2017). On the number of bipolar boolean functions. *Journal of Logic and Computation*, 27(8):2431–2449.
- [Baumann and Ulbricht, 2018] Baumann, R. and Ulbricht, M. (2018). If nothing is accepted - repairing argumentation frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018.*, pages 108–117.
- [Baumann and Woltran, 2016] Baumann, R. and Woltran, S. (2016). The role of self-attacking arguments in characterizations of equivalence notions. *Journal of Logic and Computation*, 26(4):1293–1313.
- [BBC, 2017] BBC (2017). Google’s ‘superhuman’ DeepMind AI claims chess crown. <https://www.bbc.com/news/technology-42251535>.
- [Belardinelli et al., 2015] Belardinelli, F., Grossi, D., and Maudet, N. (2015). Formal analysis of dialogues on infinite argumentation frameworks. In *IJCAI, Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, July 25-31, 2015*, pages 861–867.
- [Bench-Capon and Dunne, 2007] Bench-Capon, T. J. M. and Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641.
- [Besnard and Hunter, 2001] Besnard, P. and Hunter, A. (2001). A logic-based theory of deductive arguments. *Artificial Intelligence*, 128:203–235.
- [Besnard and Hunter, 2009] Besnard, P. and Hunter, A. (2009). Argumentation based on classical logic. In *Argumentation in Artificial Intelligence*, pages 133–152. Springer.
- [Birkhoff, 1973] Birkhoff, G. (1973). *Lattice Theory*, volume 25 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, Rhode Island, third edition.
- [Brewka, 1989] Brewka, G. (1989). Nonmonotonic logics - A brief overview. *AI Communications*, 2(2):88–97.
- [Brewka et al., 2013a] Brewka, G., Ellmauthaler, S., Strass, H., Wallner, J. P., and Woltran, S. (2013a). Abstract dialectical frameworks revisited. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pages 803–809. IJCAI/AAAI.
- [Brewka et al., 2013b] Brewka, G., Strass, H., Ellmauthaler, S., Wallner, J. P., and Woltran, S. (2013b). Abstract dialectical frameworks revisited. In *IJCAI, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pages 803–809.
- [Brewka and Woltran, 2010a] Brewka, G. and Woltran, S. (2010a). Abstract dialectical frameworks. In *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR)*, pages 102–111.
- [Brewka and Woltran, 2010b] Brewka, G. and Woltran, S. (2010b). Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010*.

- [Cabalar and Diéguez, 2014] Cabalar, P. and Diéguez, M. (2014). Strong equivalence of non-monotonic temporal theories. In Baral, C., Giacomo, G. D., and Eiter, T., editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press.
- [Caminada, 2006] Caminada, M. (2006). Semi-stable semantics. In *Computational Models of Argument: Proceedings of COMMA 2006, September 11-12, 2006, Liverpool, UK*, pages 121–130.
- [Caminada, 2011] Caminada, M. (2011). A labelling approach for ideal and stage semantics. *Argument and Computation*, 2:1–21.
- [Caminada, 2014] Caminada, M. (2014). Strong admissibility revisited. In *Computational Models of Argument - Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, pages 197–208.
- [Caminada and Amgoud, 2007] Caminada, M. and Amgoud, L. (2007). On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171:286–310.
- [Caminada et al., 2012] Caminada, M. W., Carnielli, W. A., and Dunne, P. E. (2012). Semi-stable semantics. *Journal of Logic and Computation*, 22:1207–1254.
- [Caminada and Verheij, 2010] Caminada, M. W. and Verheij, B. (2010). On the existence of semi-stable extensions. In *Benelux Conference on Artificial Intelligence*.
- [Caminada and Oren, 2014] Caminada, M. W. A. and Oren, N. (2014). Grounded semantics and infinitary argumentation frameworks. In *Benelux Conference on Artificial Intelligence*.
- [Clark, 1978] Clark, K. L. (1978). Negation as failure. In Gallaire, H. and Minker, J., editors, *Logic and Data Bases*, pages 293–322. Plenum Press.
- [Cohn, 1981] Cohn, P. M. (1981). *Universal algebra*. Mathematics and Its Applications. Springer, Dordrecht.
- [Dantsin et al., 1997] Dantsin, E., Eiter, T., Gottlob, G., and Voronkov, A. (1997). Complexity and expressive power of logic programming. In *Proceedings of the Twelfth Annual IEEE Conference on Computational Complexity, Ulm, Germany, June 24-27, 1997*, pages 82–101.
- [Davey and Priestley, 2002] Davey, B. and Priestley, H. (2002). *Introduction to Lattices and Order*. Cambridge University Press, second edition.
- [Delgrande and Peppas, 2015] Delgrande, J. P. and Peppas, P. (2015). Belief revision in horn theories. *Artificial Intelligence*, 218:1–22.
- [Delgrande et al., 2013] Delgrande, J. P., Peppas, P., and Woltran, S. (2013). Agm-style belief revision of logic programs under answer set semantics. In *Logic Programming and Nonmonotonic Reasoning, 12th International Conference, LPNMR 2013, Corunna, Spain, September 15-19, 2013. Proceedings*, pages 264–276.
- [Delgrande et al., 2008] Delgrande, J. P., Schaub, T., Tompits, H., and Woltran, S. (2008). Belief revision of logic programs under answer set semantics. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*, pages 411–421.

- [Denecker et al., 2000] Denecker, M., Marek, V., and Truszczyński, M. (2000). Approximations, Stable Operators, Well-Founded Fixpoints and Applications in Nonmonotonic Reasoning. In *Logic-Based Artificial Intelligence*, pages 127–144. Kluwer Academic Publishers.
- [Diestel, 2006] Diestel, R. (2006). *Graph Theory*. Electronic library of mathematics. Springer.
- [Diller et al., 2015] Diller, M., Haret, A., Linsbichler, T., Rümmele, S., and Woltran, S. (2015). An extension-based approach to belief revision in abstract argumentation. In *IJCAI, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 2926–2932.
- [Dung, 1995] Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357.
- [Dunne et al., 2015] Dunne, P. E., Dvořák, W., Linsbichler, T., and Woltran, S. (2015). Characteristics of multiple viewpoints in abstract argumentation. *Artificial Intelligence*, 228:153–178.
- [Dunne et al., 2013] Dunne, P. E., Dvořák, W., Linsbichler, T., and Woltran, S. (2013). Characteristics of multiple viewpoints in abstract argumentation. In *4th Workshop on Dynamics of Knowledge and Belief*, pages 16–30.
- [Dunne et al., 2016] Dunne, P. E., Spanring, C., Linsbichler, T., and Woltran, S. (2016). Investigating the relationship between argumentation semantics via signatures. In *IJCAI, Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016, New York, NY, USA, 9-15 July 2016*, pages 1051–1057.
- [Dvořák and Dunne, 2018] Dvořák, W. and Dunne, P. E. (2018). Computational problems in formal argumentation and their complexity. In Baroni, P., Gabbay, D., Giacomin, M., and van der Torre, L., editors, *Handbook of Formal Argumentation*, chapter 14. College Publications. also appears in *IfCoLog Journal of Logics and their Applications* 4(8):2557–2622.
- [Dvořák et al., 2011] Dvořák, W., Dunne, P. E., and Woltran, S. (2011). Parametric properties of ideal semantics. In *IJCAI, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 851–856.
- [Dvořák et al., 2014] Dvořák, W., Linsbichler, T., Oikarinen, E., and Woltran, S. (2014). Resolution-based grounded semantics revisited. In *Computational Models of Argument - Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, pages 269–280.
- [Dvořák and Spanring, 2017] Dvořák, W. and Spanring, C. (2017). Comparing the expressiveness of argumentation semantics. *Journal of Logic and Computation*, 27(5):1489–1521.
- [Dvořák and Gaggl, 2012] Dvořák, W. and Gaggl, S. A. (2012). Incorporating stage semantics in the scc-recursive schema for argumentation semantics. In *Proceedings of the 14th International Workshop on Non-Monotonic Reasoning (NMR 2012)*.
- [Dvořák and Woltran, 2011] Dvořák, W. and Woltran, S. (2011). On the intertranslatability of argumentation semantics. *Journal of Artificial Intelligence Research*, 41:445–475.
- [Dyrkolbotn, 2014] Dyrkolbotn, S. K. (2014). How to argue for anything: Enforcing arbitrary sets of labellings using afs. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*.

- [Eiter and Fink, 2003] Eiter, T. and Fink, M. (2003). Uniform equivalence of logic programs under the stable model semantics. In Palamidessi, C., editor, *Logic Programming, 19th International Conference, ICLP 2003, Mumbai, India, December 9-13, 2003, Proceedings*, volume 2916 of *Lecture Notes in Computer Science*, pages 224–238. Springer.
- [Eiter et al., 2013] Eiter, T., Fink, M., Pührer, J., Tompits, H., and Woltran, S. (2013). Model-based recasting in answer-set programming. *Journal of Applied Non-Classical Logics*, 23(1-2):75–104.
- [Ferrucci et al., 2010] Ferrucci, D. A., Brown, E. W., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J. M., Schlaefter, N., and Welty, C. A. (2010). Building Watson: An Overview of the DeepQA Project. *AI Magazine*, 31(3):59–79.
- [Gabbay, 1985] Gabbay, D. M. (1985). Theoretical foundations for non-monotonic reasoning in expert systems. In Apt, K. R., editor, *Logics and Models of Concurrent Systems*, pages 439–457, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Gabbay et al., 1994] Gabbay, D. M., Hogger, C. J., and Robinson, J. A., editors (1994). *Handbook of Logic in Artificial Intelligence and Logic Programming: Nonmonotonic Reasoning and Uncertain Reasoning*, volume 3. Oxford University Press, Inc., New York, NY, USA.
- [Gaggl and Dvořák, 2016] Gaggl, S. A. and Dvořák, W. (2016). Stage semantics and the SCC-recursive schema for argumentation semantics. *Journal of Logic and Computation*, 26(4):1149–1202.
- [Gaggl and Woltran, 2013] Gaggl, S. A. and Woltran, S. (2013). The cf2 argumentation semantics revisited. *Journal of Logic and Computation*, 23:925–949.
- [Gebser et al., 2012] Gebser, M., Kaminski, R., Kaufmann, B., and Schaub, T. (2012). Answer set solving in practice. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(3):1–238.
- [Gelfond and Lifschitz, 1988] Gelfond, M. and Lifschitz, V. (1988). The stable model semantics for logic programming. In *Logic Programming, Proceedings of the Fifth International Conference and Symposium, Seattle, Washington, USA, August 15-19, 1988 (2 Volumes)*, pages 1070–1080.
- [Goguen and Burstall, 1984] Goguen, J. A. and Burstall, R. M. (1984). Introducing institutions. In *Logics of Programs*, pages 221–256. Springer.
- [Gottlob, 1992] Gottlob, G. (1992). Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2(3):397–425.
- [Grossi and Modgil, 2015] Grossi, D. and Modgil, S. (2015). On the graded acceptability of arguments. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 868–874.
- [Holzinger, 2018] Holzinger, A. (2018). "explainable ai (ex-ai)". <https://gi.de/informatiklexikon/explainable-ai-ex-ai/>.
- [Jakobovits and Vermeir, 1999] Jakobovits, H. and Vermeir, D. (1999). Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9(2):215–261.
- [Kraus et al., 1990] Kraus, S., Lehmann, D., and Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1):167 – 207.
- [Lifschitz et al., 2001] Lifschitz, V., Pearce, D., and Valverde, A. (2001). Strongly equivalent logic programs. *ACM Transactions on Computational Logic*, 2(4):526–541.

- [Lifschitz and Turner, 1994] Lifschitz, V. and Turner, H. (1994). Splitting a logic program. In Hentenryck, P. V., editor, *International Conference on Logic Programming*, pages 23–37. MIT Press.
- [Linsbichler et al., 2016] Linsbichler, T., Pührer, J., and Strass, H. (2016). A uniform account of realizability in abstract argumentation. In *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*, pages 252–260.
- [Linsbichler et al., 2015] Linsbichler, T., Spanring, C., and Woltran, S. (2015). The hidden power of abstract argumentation semantics. In *Theory and Applications of Formal Argumentation - Third International Workshop, TAFA 2015, Buenos Aires, Argentina, July 25-26, 2015, Revised Selected Papers*, pages 146–162.
- [Maher, 1986] Maher, M. J. (1986). Equivalences of logic programs. In *Third International Conference on Logic Programming, Imperial College of Science and Technology, London, United Kingdom, July 14-18, 1986, Proceedings*, pages 410–424.
- [Makinson, 1994] Makinson, D. (1994). Handbook of logic in artificial intelligence and logic programming (vol. 3). chapter General Patterns in Nonmonotonic Reasoning, pages 35–110. Oxford University Press, Inc., New York, NY, USA.
- [McCarthy et al., 1955] McCarthy, J., Minsky, M., Rochester, N., and Shannon, C. (1955). Proposal for the dartmouth summer research project on artificial intelligence. <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>.
- [Nieves et al., 2011] Nieves, J. C., Osorio, M., and Zepeda, C. (2011). A schema for generating relevant logic programming semantics and its applications in argumentation theory. *Fundamenta Informaticae*, 106(2-4):295–319.
- [Oikarinen and Woltran, 2011] Oikarinen, E. and Woltran, S. (2011). Characterizing strong equivalence for argumentation frameworks. *Artificial Intelligence*, 175:1985–2009.
- [Ore, 1944] Ore, O. (1944). Galois connexions. *Transactions of the American Mathematical Society*, 55(3):493–513.
- [Osorio et al., 2005] Osorio, M., Zepeda, C., Nieves, J. C., and Cortés, U. (2005). Inferring acceptable arguments with answer set programming. In *Proceedings of the Sixth Mexican International Conference on Computer Science (ENC)*, pages 198–205.
- [Prakken, 2010] Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124.
- [Qi and Yang, 2008] Qi, G. and Yang, F. (2008). A survey of revision approaches in description logics. In *Proceedings of the 21st International Workshop on Description Logics (DL2008), Dresden, Germany, May 13-16, 2008*.
- [Rahwan and Simari, 2009] Rahwan, I. and Simari, G. R. (2009). *Argumentation in Artificial Intelligence*. Springer.
- [Reiter, 1980] Reiter, R. (1980). A Logic for Default Reasoning. *Artificial Intelligence*, 13:81–132.
- [Rudin, 1976] Rudin, W. (1976). *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill.

- [Scherer and Lotito, 2018] Scherer, M. and Lotito, M. (2018). Germany Announces 3 Billion "Artificial Intelligence (AI) Made in Germany" Initiative. <https://www.littler.com/publication-press/publication/germany-announces-eu3-billion-artificial-intelligence-ai-made-germany>.
- [Silver et al., 2018] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., and Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419):1140–1144.
- [Spanring, 2012] Spanring, C. (2012). Intertranslatability results for abstract argumentation semantics. Master's thesis, Uni Wien.
- [Spanring, 2015] Spanring, C. (2015). Hunt for the collapse of semantics in infinite abstract argumentation frameworks. In *2015 Imperial College Computing Student Workshop, ICCSW 2015, September 24-25, 2015, London, United Kingdom*, pages 70–77.
- [Strass, 2013] Strass, H. (2013). Approximating operators and semantics for abstract dialectical frameworks. *Artificial Intelligence*, 205:39–70.
- [Strass, 2015] Strass, H. (2015). The relative expressiveness of abstract argumentation and logic programming. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 1625–1631.
- [Tarski, 1955] Tarski, A. (1955). A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309.
- [Truszczyński, 2006] Truszczyński, M. (2006). Strong and uniform equivalence of nonmonotonic theories – an algebraic approach. *Annals of Mathematics and Artificial Intelligence*, 48(3–4):245–265.
- [Truszczyński, 2007] Truszczyński, M. (2007). The modal logic S4F, the default logic, and the logic here-and-there. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22–26, 2007, Vancouver, British Columbia, Canada*, pages 508–514. AAAI Press.
- [Truszczyński and Woltran, 2008] Truszczyński, M. and Woltran, S. (2008). Hyperequivalence of logic programs with respect to supported models. In Fox, D. and Gomes, C. P., editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13–17, 2008*, pages 560–565. AAAI Press.
- [Turner, 1996] Turner, H. (1996). Splitting a default theory. In Clancey, W. J. and Weld, D. S., editors, *Conference on Artificial Intelligence and Innovative Applications of Artificial Intelligence Conference*, pages 645–651. AAAI Press / The MIT Press.
- [Turner, 2001] Turner, H. (2001). Strong equivalence for logic programs and default theories (made easy). In Eiter, T., Faber, W., and Truszczyński, M., editors, *Logic Programming and Nonmonotonic Reasoning, 6th International Conference, LPNMR 2001, Vienna, Austria, September 17-19, 2001, Proceedings*, volume 2173 of *Lecture Notes in Computer Science*, pages 81–92. Springer.
- [Turner, 2004] Turner, H. (2004). Strong equivalence for causal theories. In *Logic Programming and Nonmonotonic Reasoning, 7th International Conference, LPNMR 2004, Fort Lauderdale, FL, USA, January 6-8, 2004, Proceedings*, pages 289–301.
- [van Emden and Kowalski, 1976] van Emden, M. H. and Kowalski, R. A. (1976). The Semantics of Predicate Logic as a Programming Language. *Journal of the ACM*, 23(4):733–742.

- [Verheij, 1996] Verheij, B. (1996). Two approaches to dialectical argumentation: Admissible sets and argumentation stages. In *Proceedings of the biannual International Conference on Formal and Applied Practical Reasoning (FAPR)*, pages 357–368. Universiteit.
- [Verheij, 2003] Verheij, B. (2003). Deflog: On the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13(3):319–346.
- [Weydert, 2011] Weydert, E. (2011). Semi-stable extensions for infinite frameworks. In *Benelux Conference on Artificial Intelligence*, pages 336–343.
- [Williams and Antoniou, 1998] Williams, M. and Antoniou, G. (1998). A strategy for revising default theory Extensions. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference, KR 1998, Trento, Italy, June 2-5, 1998.*, pages 24–35.
- [Woltran, 2008] Woltran, S. (2008). A common view on strong, uniform, and other notions of equivalence in answer-set programming. *Theory and Practice of Logic Programming*, 8(2):217–234.
- [Wu et al., 2009] Wu, Y., Caminada, M., and Gabbay, D. M. (2009). Complete Extensions in Argumentation Coincide with 3-Valued Stable Models in Logic Programming. *Studia Logica*, 93(2–3):383–403.
- [Zorn, 1935] Zorn, M. (1935). A remark on method in transfinite algebra. *Bulletin of the American Mathematical Society*, 41:667–670.

Selbstständigkeitserklärung

Hiermit erkläre ich, die vorliegende Habilitationsschrift selbstständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

.....
(Ort, Datum)

.....
(Unterschrift)

Erklärung zu Kenntnis und Anerkennung der gültigen Habilitationsordnung

Hiermit erkläre ich, die zum Zeitpunkt der Stellung dieses Habilitationsantrages gültige Habilitationsordnung der Fakultät für Mathematik und Informatik der Universität Leipzig (aktuell: vom 9. Juli 1998) zu kennen und anzuerkennen.

.....
(Ort, Datum)

.....
(Unterschrift)